

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
13 March 2003 (13.03.2003)

PCT

(10) International Publication Number
WO 03/020953 A2

- (51) International Patent Classification⁷: **C12Q**
- (21) International Application Number: PCT/US02/27728
- (22) International Filing Date: 29 August 2002 (29.08.2002)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/316,259 31 August 2001 (31.08.2001) US
- (63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/316,259 (CIP)
Filed on 31 August 2001 (31.08.2001)
- (71) Applicant (for all designated States except US): **DI-ADEXUS, INC.** [US/US]; 343 Oyster Point Boulevard, South San Francisco, CA 94080 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **SUN, Yongming** [CN/US]; 551 Shoal Drive, Redwood City, CA 94065 (US). **LIU, Chenghua** [CN/US]; 1125 Ranchero Way #14, San Jose, CA 95117 (US). **GHOSH, Malavika, G.** [IN/US]; 1733 Via Lugano, San Jose, CA 95120 (US).
- (74) Agents: **LICATA, Jane, Massey** et al.; Licata & Tyrrell P.C., 66 E. Main Street, Marlton, NJ 08053 (US).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SI, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: COMPOSITIONS AND METHODS RELATING TO COLON SPECIFIC GENES AND PROTEINS

(57) Abstract: The present invention relates to newly identified nucleic acid molecules and polypeptides present in normal and neoplastic colon cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions containing the nucleic acid molecules, polypeptides, antibodies, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating colon cancer and non-cancerous disease states in colon, identifying colon tissue, monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered colon tissue for treatment and research.



WO 03/020953 A2

COMPOSITIONS AND METHODS RELATING TO COLON SPECIFIC GENES AND PROTEINS

INTRODUCTION

5 This application claims the benefit of priority from U.S. Provisional Application No. 60/316,259, filed August 31, 2001, which is herein incorporated by reference in its entirety.

FIELD OF THE INVENTION

10 The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic colorectal cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of
15 the invention and methods for the use of these compositions. This invention relates to newly developed assays for detecting, diagnosing, monitoring, staging, prognosticating, imaging and treating cancers, particularly colorectal cancer, and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered
20 colorectal tissue for treatment and research.

BACKGROUND OF THE INVENTION

 Colorectal cancer is the second most common cause of cancer death in the United States and the third most prevalent cancer in both men and women. M. L. Davila & A. D.
25 Davila, *Screening for Colon and Rectal Cancer*, in Colon and Rectal Cancer 47 (Peter S. Edelstein ed., 2000). Approximately 100,000 patients every year suffer from colon cancer and approximately half that number die of the disease. Hannah-Ngoc Ha & Bard C. Cosman, *Treatment of Colon Cancer*, in Colon and Rectal Cancer 157 (Peter S. Edelstein ed., 2000). Nearly all cases of colorectal cancer arise from adenomatous polyps, some of
30 which mature into large polyps, undergo abnormal growth and development, and ultimately progress into cancer. Davila & Davila, *supra* at 55-56. This progression would appear to take at least 10 years in most patients, rendering it a readily treatable form of

cancer if diagnosed early, when the cancer is localized. *Id.* at 56; Walter J. Burdette, Cancer: Etiology, Diagnosis, and Treatment 125 (1998).

Although our understanding of the etiology of colon cancer is undergoing continual refinement, extensive research in this area points to a combination of factors, including age, hereditary and nonhereditary conditions, and environmental/dietary factors. Age is a key risk factor in the development of colorectal cancer, Davila & Davila, *supra* at 48, with men and women over 40 years of age become increasingly susceptible to that cancer, Burdette, *supra* at 126. Incidence rates increase considerably in each subsequent decade of life. Davila et al., *supra* at 48. A number of hereditary and nonhereditary conditions have also been linked to a heightened risk of developing colorectal cancer, including familial adenomatous polyposis (FAP), hereditary nonpolyposis colorectal cancer (Lynch syndrome or HNPCC), a personal and/or family history of colorectal cancer or adenomatous polyps, inflammatory bowel disease, diabetes mellitus, and obesity. *Id.* at 47; Henry T. Lynch & Jane F. Lynch, *Hereditary Nonpolyposis Colorectal Cancer (Lynch Syndromes)*, in Colon and Rectal Cancer 67-68 (Peter S. Edelstein ed., 2000). In the case of FAP, the tumor suppressor gene APC (adenomatous polyposis coli), located at 5q21, has been either mutationally inactivated or deleted. Alberts et al., Molecular Biology of the Cell 1288 (3d ed. 1994). The APC protein plays a role in a number of functions, including cell adhesion, apoptosis, and repression of the *c-myc* oncogene. N. R. Hall & R. D. Madoff, *Genetics and the Polyp-Cancer Sequence*, Colon and Rectal Cancer 8 (Peter S. Edelstein, ed., 2000). Of those patients with colorectal cancer who have normal APC genes, over 65% have such mutations in the cancer cells but not in other tissues. Alberts et al., *supra* at 1288. In the case of HNPCC, patients manifest abnormalities in the tumor suppressor gene HNPCC, but only about 15% of tumors contain the mutated gene. *Id.* A host of other genes have also been implicated in colorectal cancer, including the *K-ras*, *N-ras*, *H-ras* and *c-myc* oncogenes, and the tumor suppressor genes *DCC* (deleted in colon carcinoma) and *p53*. Hall & Madoff, *supra* at 8-9; Alberts et al., *supra* at 1288.

Environmental/dietary factors associated with an increased risk of colorectal cancer include a high fat diet, intake of high dietary red meat, and sedentary lifestyle. Davila & Davila, *supra* at 47; Reddy, B. S., *Prev. Med.* 16(4): 460-7 (1987). Conversely, environmental/dietary factors associated with a reduced risk of colorectal cancer include a diet high in fiber, folic acid, calcium, and hormone-replacement therapy in post-

menopausal women. Davila & Davila, *supra* at 50-55. The effect of antioxidants in reducing the risk of colon cancer is unclear. *Id.* at 53.

Because colon cancer is highly treatable when detected at an early, localized stage, screening should be a part of routine care for all adults starting at age 50, especially those with first-degree relatives with colorectal cancer. One major advantage of colorectal cancer screening over its counterparts in other types of cancer is its ability to not only detect precancerous lesions, but to remove them as well. Davila & Davila, *supra* at 56. The key colorectal cancer screening tests in use today are fecal occult blood test, sigmoidoscopy, colonoscopy, double-contrast barium enema, and the carcinoembryonic antigen (CEA) test. *Id.*; Burdette, *supra* at 125.

The fecal occult blood test (FOBT) screens for colorectal cancer by detecting the amount of blood in the stool, the premise being that neoplastic tissue, particularly malignant tissue, bleeds more than typical mucosa, with the amount of bleeding increasing with polyp size and cancer stage. Davila & Davila, *supra* at 56-57. While effective at detecting early stage tumors, FOBT is unable to detect adenomatous polyps (premalignant lesions), and, depending on the contents of the fecal sample, is subject to rendering false positives. *Id.* at 56-59. Sigmoidoscopy and colonoscopy, by contrast, allow direct visualization of the bowel, and enable one to detect, biopsy, and remove adenomatous polyps. *Id.* at 59-60, 61. Despite the advantages of these procedures, there are accompanying downsides: sigmoidoscopy, by definition, is limited to the sigmoid colon and below, colonoscopy is a relatively expensive procedure, and both share the risk of possible bowel perforation and hemorrhaging. *Id.* at 59-60. Double-contrast barium enema (DCBE) enables detection of lesions better than FOBT, and almost as well as colonoscopy, but it may be limited in evaluating the winding rectosigmoid region. *Id.* at 60. The CEA blood test, which involves screening the blood for carcinoembryonic antigen, shares the downside of FOBT, in that it is of limited utility in detecting colorectal cancer at an early stage. Burdette, *supra* at 125.

Once colon cancer has been diagnosed, treatment decisions are typically made in reference to the stage of cancer progression. A number of techniques are employed to stage the cancer (some of which are also used to screen for colon cancer), including pathologic examination of resected colon, sigmoidoscopy, colonoscopy, and various imaging techniques. AJCC Cancer Staging Handbook 84 (Irvin D. Fleming et al. eds., 5th ed. 1998); Montgomery, R. C. and Ridge, J.A., *Semin. Surg. Oncol.* 15(3): 143-150

(1998). Moreover, chest films, liver functionality tests, and liver scans are employed to determine the extent of metastasis. Fleming et al. eds., *supra* at 84. While computerized tomography and magnetic resonance imaging are useful in staging colorectal cancer in its later stages, both have unacceptably low staging accuracy for identifying early stages of the disease, due to the difficulty that both methods have in (1) revealing the depth of bowel wall tumor infiltration and (2) diagnosing malignant adenopathy. Thoeni, R. F., *Radiol. Clin. N. Am.* 35(2): 457-85 (1997). Rather, techniques such as transrectal ultrasound (TRUS) are preferred in this context, although this technique is inaccurate with respect to detecting small lymph nodes that may contain metastases. David Blumberg & Frank G. Opelka, *Neoadjuvant and Adjuvant Therapy for Adenocarcinoma of the Rectum, in Colon and Rectal Cancer* 316 (Peter S. Edelstein ed., 2000).

Several classification systems have been devised to stage the extent of colorectal cancer, including the Dukes' system and the more detailed International Union against Cancer-American Joint Committee on Cancer TNM staging system, which is considered by many in the field to be a more useful staging system. Burdette, *supra* at 126-27. The TNM system, which is used for either clinical or pathological staging, is divided into four stages, each of which evaluates the extent of cancer growth with respect to primary tumor (T), regional lymph nodes (N), and distant metastasis (M). Fleming et al. eds., *supra* at 84-85. The system focuses on the extent of tumor invasion into the intestinal wall, invasion of adjacent structures, the number of regional lymph nodes that have been affected, and whether distant metastasis has occurred. *Id.* at 81.

Stage 0 is characterized by *in situ* carcinoma (Tis), in which the cancer cells are located inside the glandular basement membrane (intraepithelial) or lamina propria (intramucosal). *Id.* at 84-85; Burdette, *supra* at 127. In this stage, the cancer has not spread to the regional lymph nodes (N0), and there is no distant metastasis (M0). Fleming et al. eds., *supra* at 85; Burdette, *supra* at 127. In stage I, there is still no spread of the cancer to the regional lymph nodes and no distant metastasis, but the tumor has invaded the submucosa (T1) or has progressed further to invade the muscularis propria (T2). Fleming et al. eds., *supra* at 84-85; Burdette, *supra* at 127. Stage II also involves no spread of the cancer to the regional lymph nodes and no distant metastasis, but the tumor has invaded the subserosa, or the nonperitonealized pericolic or perirectal tissues (T3), or has progressed to invade other organs or structures, and/or has perforated the visceral peritoneum (T4). *Id.* Stage 3 is characterized by any of the T substages, no distant

metastasis, and either metastasis in 1 to 3 regional lymph nodes (N1) or metastasis in four or more regional lymph nodes (N2). Fleming et al. eds., *supra* at 85; Burdette, *supra* at 127. Lastly, stage 4 involves any of the T or N substages, as well as distant metastasis. *Id.*

5 Currently, pathological staging of colon cancer is preferable over clinical staging as pathological staging provides a more accurate prognosis. Pathological staging typically involves examination of the resected colon section, along with surgical examination of the abdominal cavity. Fleming et al. eds., *supra* at 84. Clinical staging would be a preferred method of staging were it at least as accurate as pathological staging, as it does not
10 depend on the invasive procedures of its counterpart.

Turning to the treatment of colorectal cancer, surgical resection results in a cure for roughly 50% of patients. Burdette, *supra* at 125. Irradiation is used both preoperatively and postoperatively in treating colorectal cancer. *Id.* at 125, 132-33. Chemotherapeutic agents, particularly 5-fluorouracil, are also powerful weapons in treating colorectal cancer.
15 *Id.* at 125, 133. Other agents include irinotecan and floxuridine, cisplatin, levamisole, methotrexate, interferon- α , and leucovorin. *Id.* at 133. Nonetheless, thirty to forty percent of patients will develop a recurrence of colon cancer following surgical resection. Wayne De Vos, *Follow-up After Treatment of Colon Cancer, Colon and Rectal Cancer* 225 (Peter S. Edelstein ed., 2000), which in many patients is the ultimate cause of death.

20 Accordingly, colon cancer patients must be closely monitored to determine response to therapy and to detect persistent or recurrent disease and metastasis. From the foregoing, it is clear that procedures used for detecting, diagnosing, monitoring, staging, prognosticating, and preventing the recurrence of colorectal cancer are of critical importance to the outcome of the patient. Moreover, current procedures, while helpful in
25 each of these analyses, are limited by their specificity, sensitivity, invasiveness, and/or their cost. As such, highly specific and sensitive procedures that would operate by way of detecting novel markers in cells, tissues, or bodily fluids, with minimal invasiveness and at a reasonable cost, would be highly desirable.

Accordingly, there is a great need for more sensitive and accurate methods for
30 predicting whether a person is likely to develop colorectal cancer, for diagnosing colorectal cancer, for monitoring the progression of the disease, for staging the colorectal cancer, for determining whether the colorectal cancer has metastasized, and for imaging the colorectal cancer. There is also a need for better treatment of colorectal cancer.

SUMMARY OF THE INVENTION

The present invention solves many needs in the art by providing nucleic acid molecules, polypeptides and antibodies thereto, variants and derivatives of the nucleic acids and polypeptides, agonists and antagonists that may be used to identify, diagnose, monitor, stage, image and treat colon cancer and non-cancerous disease states in colon; identify and monitor colon tissue; and identify and design agonists and antagonists of polypeptides of the invention. The invention also provides gene therapy, methods for producing transgenic animals and cells, and methods for producing engineered colon tissue for treatment and research.

One aspect of the present invention relates to nucleic acid molecules that are specific to colon cells, colon tissue and/or the colon organ. These colon specific nucleic acids (CSNAs) may be a naturally-occurring cDNA, genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. If the CSNA is genomic DNA, then the CSNA is a colon specific gene (CSG). In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to colon. More preferred is a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence of SEQ ID NO: 101-190. In another preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1-100. For the sequences listed herein, DEX0234_1 corresponds to SEQ ID NO: 1, DEX0234_101 corresponds to SEQ ID NO: 101, etc.

This aspect of the present invention also relates to nucleic acid molecules that selectively hybridize or exhibit substantial sequence similarity to nucleic acid molecules encoding a Colon Specific Protein (CSP), or that selectively hybridize or exhibit substantial sequence similarity to a CSNA. In one embodiment of the present invention the nucleic acid molecule comprises an allelic variant of a nucleic acid molecule encoding a CSP, or an allelic variant of a CSNA. In another embodiment, the nucleic acid molecule comprises a part of a nucleic acid sequence that encodes a CSP or a part of a nucleic acid sequence of a CSNA.

In addition, this aspect of the present invention relates to a nucleic acid molecule further comprising one or more expression control sequences controlling the transcription and/or translation of all or a part of a CSNA or the the transcription and/or translation of a nucleic acid molecule that encodes all or a fragment of a CSP.

Another aspect of the present invention relates to vectors and/or host cells comprising a nucleic acid molecule of this invention. In a preferred embodiment, the nucleic acid molecule of the vector and/or host cell encodes all or a fragment of a CSP. In another preferred embodiment, the nucleic acid molecule of the vector and/or host cell
5 comprises all or a part of a CSNA. Vectors and host cells of the present invention are useful in the recombinant production of polypeptides, particularly CSPs of the present invention.

Another aspect of the present invention relates to polypeptides encoded by a nucleic acid molecule of this invention. The polypeptide may comprise either a fragment
10 or a full-length protein. In a preferred embodiment, the polypeptide is a CSP. However, this aspect of the present invention also relates to mutant proteins (muteins) of CSPs, fusion proteins of which a portion is a CSP, and proteins and polypeptides encoded by allelic variants of a CSNA as provided herein.

Another aspect of the present invention relates to antibodies and other binders that
15 specifically binds to a polypeptide of the instant invention. Accordingly antibodies or binders of the present specifically bind to CSPs, muteins, fusion proteins, and/or homologous proteins or a polypeptides encoded by allelic variants of an CSNA as provided herein.

Another aspect of the present invention relates to agonists and antagonists of the
20 nucleic acid molecules and polypeptides of this invention. The agonists and antagonists of the instant invention may be used to treat colon cancer and non-cancerous disease states in colon and to produce engineered colon tissue.

Another aspect of the present invention relates to methods for using the nucleic acid molecules to detect or amplify nucleic acid molecules that have similar or identical
25 nucleic acid sequences compared to the nucleic acid molecules described herein. Such methods are useful in identifying, diagnosing, monitoring, staging, imaging and treating colon cancer and non-cancerous disease states in colon. Such methods are also useful in identifying and/or monitoring colon tissue. In addition, measurement of levels of the nucleic acid molecules of this invention may be useful for diagnostics as part of panel in
30 combination with other markers.

Another aspect of the present invention relates to use of the nucleic acid molecules of this invention in gene therapy, for producing transgenic animals and cells, and for producing engineered colon tissue for treatment and research.

Another aspect of the present invention relates to methods for detecting polypeptides this invention, preferably using antibodies thereto. Such methods are useful to identify, diagnose, monitor, stage, image and treat colon cancer and non-cancerous disease states in colon. In addition, measurement of levels of the polypeptides of this invention may be useful to identify, diagnose, monitor, stage, image colon cancer in combination with other colon cancer markers. The polypeptides of the present invention can also be used to identify and/or monitor colon tissue, and to produce engineered colon tissue.

Yet another aspect of the present invention relates to a computer readable means of storing the nucleic acid and amino acid sequences of the invention. The records of the computer readable means can be accessed for reading and displaying of sequences for comparison, alignment and ordering of the sequences of the invention to other sequences. In addition, the computer records regarding the nucleic acid and/or amino acid sequences and/or measurements of their levels may be used alone or in combination with other markers to diagnose colon related diseases.

DETAILED DESCRIPTION OF THE INVENTION

Definitions and General Techniques

Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well-known and commonly used in the art. The methods and techniques of the present invention are generally performed according to conventional methods well-known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. *See, e.g.,* Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press (1989) and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 3d ed., Cold Spring Harbor Press (2001); Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing Associates (1992, and Supplements to 2000); Ausubel *et al.*, Short Protocols in Molecular Biology: A

Compendium of Methods from Current Protocols in Molecular Biology – 4th Ed., Wiley & Sons (1999); Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1990); and Harlow and Lane, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1999).

5 Enzymatic reactions and purification techniques are performed according to manufacturer's specifications, as commonly accomplished in the art or as described herein. The nomenclatures used in connection with, and the laboratory procedures and techniques of, analytical chemistry, synthetic organic chemistry, and medicinal and pharmaceutical chemistry described herein are those well-known and commonly used in
10 the art. Standard techniques are used for chemical syntheses, chemical analyses, pharmaceutical preparation, formulation, and delivery, and treatment of patients.

The following terms, unless otherwise indicated, shall be understood to have the following meanings:

A "nucleic acid molecule" of this invention refers to a polymeric form of
15 nucleotides and includes both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. A nucleotide refers to a ribonucleotide, deoxynucleotide or a modified form of either type of nucleotide. A "nucleic acid molecule" as used herein is synonymous with "nucleic acid" and "polynucleotide." The term "nucleic acid molecule" usually refers to a molecule of at
20 least 10 bases in length, unless otherwise specified. The term includes single and double stranded forms of DNA. In addition, a polynucleotide may include either or both naturally-occurring and modified nucleotides linked together by naturally-occurring and/or non-naturally occurring nucleotide linkages.

The nucleic acid molecules may be modified chemically or biochemically or may
25 contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (*e.g.*,
30 phosphorothioates, phosphorodithioates, etc.), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine, psoralen, etc.), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids, etc.) The term "nucleic acid molecule" also includes any topological conformation, including single-stranded, double-stranded, partially

duplexed, triplexed, hairpinned, circular and padlocked conformations. Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

A "gene" is defined as a nucleic acid molecule that comprises a nucleic acid sequence that encodes a polypeptide and the expression control sequences that surround the nucleic acid sequence that encodes the polypeptide. For instance, a gene may comprise a promoter, one or more enhancers, a nucleic acid sequence that encodes a polypeptide, downstream regulatory sequences and, possibly, other nucleic acid sequences involved in regulation of the expression of an RNA. As is well-known in the art, eukaryotic genes usually contain both exons and introns. The term "exon" refers to a nucleic acid sequence found in genomic DNA that is bioinformatically predicted and/or experimentally confirmed to contribute contiguous sequence to a mature mRNA transcript. The term "intron" refers to a nucleic acid sequence found in genomic DNA that is predicted and/or confirmed to not contribute to a mature mRNA transcript, but rather to be "spliced out" during processing of the transcript.

A nucleic acid molecule or polypeptide is "derived" from a particular species if the nucleic acid molecule or polypeptide has been isolated from the particular species, or if the nucleic acid molecule or polypeptide is homologous to a nucleic acid molecule or polypeptide isolated from a particular species.

An "isolated" or "substantially pure" nucleic acid or polynucleotide (*e.g.*, an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, *e.g.*, ribosomes, polymerases, or genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a portion of a polynucleotide in which the "isolated polynucleotide" is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, (4) does not occur in nature as part of a larger sequence or (5) includes nucleotides or internucleoside bonds that are not found in nature. The term "isolated" or "substantially pure" also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous

systems. The term "isolated nucleic acid molecule" includes nucleic acid molecules that are integrated into a host cell chromosome at a heterologous site, recombinant fusions of a native fragment to a heterologous sequence, recombinant vectors present as episomes or as integrated into a host cell chromosome.

5 A "part" of a nucleic acid molecule refers to a nucleic acid molecule that comprises a partial contiguous sequence of at least 10 bases of the reference nucleic acid molecule. Preferably, a part comprises at least 15 to 20 bases of a reference nucleic acid molecule. In theory, a nucleic acid sequence of 17 nucleotides is of sufficient length to occur at random less frequently than once in the three gigabase human genome, and thus
10 to provide a nucleic acid probe that can uniquely identify the reference sequence in a nucleic acid mixture of genomic complexity. A preferred part is one that comprises a nucleic acid sequence that can encode at least 6 contiguous amino acid sequences (fragments of at least 18 nucleotides) because they are useful in directing the expression or synthesis of peptides that are useful in mapping the epitopes of the polypeptide encoded
15 by the reference nucleic acid. *See, e.g., Geysen et al., Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1984); and United States Patent Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. A part may also comprise at least 25, 30, 35 or 40 nucleotides of a reference nucleic acid molecule, or at least 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides of a
20 reference nucleic acid molecule. A part of a nucleic acid molecule may comprise no other nucleic acid sequences. Alternatively, a part of a nucleic acid may comprise other nucleic acid sequences from other nucleic acid molecules.

 The term "oligonucleotide" refers to a nucleic acid molecule generally comprising a length of 200 bases or fewer. The term often refers to single-stranded
25 deoxyribonucleotides, but it can refer as well to single-or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs, among others. Preferably, oligonucleotides are 10 to 60 bases in length and most preferably 12, 13, 14, 15, 16, 17, 18, 19 or 20 bases in length. Other preferred oligonucleotides are 25, 30, 35, 40, 45, 50, 55 or 60 bases in length. Oligonucleotides may be single-stranded, *e.g.* for use as probes
30 or primers, or may be double-stranded, *e.g.* for use in the construction of a mutant gene. Oligonucleotides of the invention can be either sense or antisense oligonucleotides. An oligonucleotide can be derivatized or modified as discussed above for nucleic acid molecules.

Oligonucleotides, such as single-stranded DNA probe oligonucleotides, often are synthesized by chemical methods, such as those implemented on automated oligonucleotide synthesizers. However, oligonucleotides can be made by a variety of other methods, including in vitro recombinant DNA-mediated techniques and by
5 expression of DNAs in cells and organisms. Initially, chemically synthesized DNAs typically are obtained without a 5' phosphate. The 5' ends of such oligonucleotides are not substrates for phosphodiester bond formation by ligation reactions that employ DNA ligases typically used to form recombinant DNA molecules. Where ligation of such oligonucleotides is desired, a phosphate can be added by standard techniques, such as
10 those that employ a kinase and ATP. The 3' end of a chemically synthesized oligonucleotide generally has a free hydroxyl group and, in the presence of a ligase, such as T4 DNA ligase, readily will form a phosphodiester bond with a 5' phosphate of another polynucleotide, such as another oligonucleotide. As is well-known, this reaction can be prevented selectively, where desired, by removing the 5' phosphates of the other
15 polynucleotide(s) prior to ligation.

The term "naturally-occurring nucleotide" referred to herein includes naturally-occurring deoxyribonucleotides and ribonucleotides. The term "modified nucleotides" referred to herein includes nucleotides with modified or substituted sugar groups and the like. The term "nucleotide linkages" referred to herein includes nucleotides linkages such
20 as phosphorothioate, phosphorodithioate, phosphoroselenoate, phosphorodiselenoate, phosphoroanilothioate, phosphoraniladate, phosphoroamidate, and the like. *See e.g.*, LaPlanche *et al. Nucl. Acids Res.* 14:9081-9093 (1986); Stein *et al. Nucl. Acids Res.* 16:3209-3221 (1988); Zon *et al. Anti-Cancer Drug Design* 6:539-568 (1991); Zon *et al.*, in Eckstein (ed.) Oligonucleotides and Analogues: A Practical Approach, pp. 87-108,
25 Oxford University Press (1991); Uhlmann and Peyman *Chemical Reviews* 90:543 (1990), and United States Patent No. 5,151,510, the disclosure of which is hereby incorporated by reference in its entirety.

Unless specified otherwise, the left hand end of a polynucleotide sequence in sense orientation is the 5' end and the right hand end of the sequence is the 3' end. In addition,
30 the left hand direction of a polynucleotide sequence in sense orientation is referred to as the 5' direction, while the right hand direction of the polynucleotide sequence is referred to as the 3' direction. Further, unless otherwise indicated, each nucleotide sequence is set forth herein as a sequence of deoxyribonucleotides. It is intended, however, that the given

sequence be interpreted as would be appropriate to the polynucleotide composition: for example, if the isolated nucleic acid is composed of RNA, the given sequence intends ribonucleotides, with uridine substituted for thymidine.

The term “allelic variant” refers to one of two or more alternative naturally-
5 occurring forms of a gene, wherein each gene possesses a unique nucleotide sequence. In a preferred embodiment, different alleles of a given gene have similar or identical biological properties.

The term “percent sequence identity” in the context of nucleic acid sequences refers to the residues in two sequences which are the same when aligned for maximum
10 correspondence. The length of sequence identity comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number of different algorithms known in the art which can be used to measure nucleotide sequence
15 identity. For instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA, which includes, *e.g.*, the programs FASTA2 and FASTA3, provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson, *Methods Enzymol.* 183: 63-98
20 (1990); Pearson, *Methods Mol. Biol.* 132: 185-219 (2000); Pearson, *Methods Enzymol.* 266: 227-258 (1996); Pearson, *J. Mol. Biol.* 276: 71-84 (1998)). Unless otherwise specified, default parameters for a particular program or algorithm are used. For instance, percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the
25 scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1.

A reference to a nucleic acid sequence encompasses its complement unless otherwise specified. Thus, a reference to a nucleic acid molecule having a particular sequence should be understood to encompass its complementary strand, with its complementary sequence. The complementary strand is also useful, *e.g.*, for antisense
30 therapy, hybridization probes and PCR primers.

In the molecular biology art, researchers use the terms “percent sequence identity”, “percent sequence similarity” and “percent sequence homology” interchangeably. In this

application, these terms shall have the same meaning with respect to nucleic acid sequences only.

The term “substantial similarity” or “substantial sequence similarity,” when referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases, as measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

Alternatively, substantial similarity exists between a first and second nucleic acid sequence when the first nucleic acid sequence or fragment thereof hybridizes to an antisense strand of the second nucleic acid, , under selective hybridization conditions. Typically, selective hybridization will occur between the first nucleic acid sequence and an antisense strand of the second nucleic acid sequence when there is at least about 55% sequence identity between the first and second nucleic acid sequences— preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90% — over a stretch of at least about 14 nucleotides, more preferably at least 17 nucleotides, even more preferably at least 20, 25, 30, 35, 40, 50, 60, 70, 80, 90 or 100 nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species, length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. “Stringent hybridization conditions” and “stringent wash conditions” in the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. The most important parameters include temperature of hybridization, base composition of the nucleic acids, salt concentration and length of the nucleic acid. One having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization. In general, “stringent hybridization” is performed at about 25°C below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of conditions. “Stringent washing” is performed at temperatures about 5°C lower than the T_m for the specific DNA hybrid under a particular set of conditions. The T_m is the

temperature at which 50% of the target sequence hybridizes to a perfectly matched probe.

See Sambrook (1989), *supra*, p. 9.51.

The T_m for a particular DNA-DNA hybrid can be estimated by the formula:

$$T_m = 81.5^\circ\text{C} + 16.6 (\log_{10}[\text{Na}^+]) + 0.41 (\text{fraction G} + \text{C}) -$$

5 $0.63 (\% \text{ formamide}) - (600/l)$ where l is the length of the hybrid in base pairs.

The T_m for a particular RNA-RNA hybrid can be estimated by the formula:

$$T_m = 79.8^\circ\text{C} + 18.5 (\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) +$$

$$11.8 (\text{fraction G} + \text{C})^2 - 0.35 (\% \text{ formamide}) - (820/l).$$

The T_m for a particular RNA-DNA hybrid can be estimated by the formula:

10 $T_m = 79.8^\circ\text{C} + 18.5(\log_{10}[\text{Na}^+]) + 0.58 (\text{fraction G} + \text{C}) +$

$$11.8 (\text{fraction G} + \text{C})^2 - 0.50 (\% \text{ formamide}) - (820/l).$$

In general, the T_m decreases by 1-1.5°C for each 1% of mismatch between two nucleic acid sequences. Thus, one having ordinary skill in the art can alter hybridization and/or washing conditions to obtain sequences that have higher or lower degrees of
 15 sequence identity to the target nucleic acid. For instance, to obtain hybridizing nucleic acids that contain up to 10% mismatch from the target nucleic acid sequence, 10-15°C would be subtracted from the calculated T_m of a perfectly matched hybrid, and then the hybridization and washing temperatures adjusted accordingly. Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other
 20 higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well-known in the art.

An example of stringent hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 50% formamide/6X SSC
 25 at 42°C for at least ten hours and preferably overnight (approximately 16 hours). Another example of stringent hybridization conditions is 6X SSC at 68°C without formamide for at least ten hours and preferably overnight. An example of moderate stringency hybridization conditions is 6X SSC at 55°C without formamide for at least ten hours and preferably overnight. An example of low stringency hybridization conditions for
 30 hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or northern blot or for screening a library is 6X SSC at 42°C for at least ten hours. Hybridization conditions to identify nucleic acid sequences that are similar but not identical can be identified by experimentally changing

the hybridization temperature from 68°C to 42°C while keeping the salt concentration constant (6X SSC), or keeping the hybridization temperature and salt concentration constant (e.g. 42°C and 6X SSC) and varying the formamide concentration from 50% to 0%. Hybridization buffers may also include blocking agents to lower background. These
5 agents are well-known in the art. *See* Sambrook *et al.* (1989), *supra*, pages 8.46 and 9.46-9.58. *See also* Ausubel (1992), *supra*, Ausubel (1999), *supra*, and Sambrook (2001), *supra*.

Wash conditions also can be altered to change stringency conditions. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see* Sambrook
10 (1989), *supra*, for SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove excess probe. An exemplary medium stringency wash for duplex DNA of more than 100 base pairs is 1x SSC at 45°C for 15 minutes. An exemplary low stringency wash for such a duplex is 4x SSC at 40°C for 15 minutes. In general, signal-to-noise ratio of 2x or higher than that observed for an unrelated probe in
15 the particular hybridization assay indicates detection of a specific hybridization.

As defined herein, nucleic acids that do not hybridize to each other under stringent conditions are still substantially similar to one another if they encode polypeptides that are substantially identical to each other. This occurs, for example, when a nucleic acid is created synthetically or recombinantly using a high codon degeneracy as permitted by the
20 redundancy of the genetic code.

Hybridization conditions for nucleic acid molecules that are shorter than 100 nucleotides in length (e.g., for oligonucleotide probes) may be calculated by the formula:
$$T_m = 81.5^{\circ}\text{C} + 16.6(\log_{10}[\text{Na}^+]) + 0.41(\text{fraction G+C}) - (600/\text{N}),$$
 wherein N is
change length and the $[\text{Na}^+]$ is 1 M or less. *See* Sambrook (1989), *supra*, p. 11.46. For
25 hybridization of probes shorter than 100 nucleotides, hybridization is usually performed under stringent conditions (5-10°C below the T_m) using high concentrations (0.1-1.0 pmol/ml) of probe. *Id.* at p. 11.45. Determination of hybridization using mismatched probes, pools of degenerate probes or “guessmers,” as well as hybridization solutions and methods for empirically determining hybridization conditions are well-known in the art.
30 *See, e.g.*, Ausubel (1999), *supra*; Sambrook (1989), *supra*, pp. 11.45-11.57.

The term “digestion” or “digestion of DNA” refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes referred to herein are commercially available and their

reaction conditions, cofactors and other requirements for use are known and routine to the skilled artisan. For analytical purposes, typically, 1 μ g of plasmid or DNA fragment is digested with about 2 units of enzyme in about 20 μ l of reaction buffer. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 μ g of DNA are

5 digested with 20 to 250 units of enzyme in proportionately larger volumes. Appropriate buffers and substrate amounts for particular restriction enzymes are described in standard laboratory manuals, such as those referenced below, and they are specified by commercial suppliers. Incubation times of about 1 hour at 37°C are ordinarily used, but conditions may vary in accordance with standard procedures, the supplier's instructions and the particulars

10 of the reaction. After digestion, reactions may be analyzed, and fragments may be purified by electrophoresis through an agarose or polyacrylamide gel, using well-known methods that are routine for those skilled in the art.

The term "ligation" refers to the process of forming phosphodiester bonds between two or more polynucleotides, which most often are double-stranded DNAs. Techniques

15 for ligation are well-known to the art and protocols for ligation are described in standard laboratory manuals and references, such as, *e.g.*, Sambrook (1989), *supra*.

Genome-derived "single exon probes," are probes that comprise at least part of an exon ("reference exon") and can hybridize detectably under high stringency conditions to transcript-derived nucleic acids that include the reference exon but do not hybridize

20 detectably under high stringency conditions to nucleic acids that lack the reference exon. Single exon probes typically further comprise, contiguous to a first end of the exon portion, a first intronic and/or intergenic sequence that is identically contiguous to the exon in the genome, and may contain a second intronic and/or intergenic sequence that is identically contiguous to the exon in the genome. The minimum length of genome-

25 derived single exon probes is defined by the requirement that the exonic portion be of sufficient length to hybridize under high stringency conditions to transcript-derived nucleic acids, as discussed above. The maximum length of genome-derived single exon probes is defined by the requirement that the probes contain portions of no more than one exon. The single exon probes may contain priming sequences not found in contiguity with

30 the rest of the probe sequence in the genome, which priming sequences are useful for PCR and other amplification-based technologies. In another aspect, the invention is directed to single exon probes based on the CSNAs disclosed herein.

In one embodiment, the term “microarray” refers to a “nucleic acid microarray” having a substrate-bound plurality of nucleic acids, hybridization to each of the plurality of bound nucleic acids being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed. Nucleic acid microarrays include all the devices so called in Schena (ed.), DNA Microarrays: A Practical Approach (Practical Approach Series), Oxford University Press (1999); *Nature Genet.* 21(1)(suppl.):1 - 60 (1999); Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing Company/BioTechniques Books Division (2000). Additionally, these nucleic acid microarrays include substrate-bound plurality of nucleic acids in which the plurality of nucleic acids are disposed on a plurality of beads, rather than on a unitary planar substrate, as is described, *inter alia*, in Brenner *et al.*, *Proc. Natl. Acad. Sci. USA* 97(4):1665-1670 (2000). Examples of nucleic acid microarrays may be found in U.S. Patent Nos. 6,391,623, 6,383,754, 6,383,749, 6,380,377, 6,379,897, 6,376,191, 6,372,431, 6,351,712, 6,344,316, 6,316,193, 6,312,906, 6,309,828, 6,309,824, 6,306,643, 6,300,063, 6,287,850, 6,284,497, 6,284,465, 6,280,954, 6,262,216, 6,251,601, 6,245,518, 6,263,287, 6,251,601, 6,238,866, 6,228,575, 6,214,587, 6,203,989, 6,171,797, 6,103,474, 6,083,726, 6,054,274, 6,040,138, 6,083,726, 6,004,755, 6,001,309, 5,958,342, 5,952,180, 5,936,731, 5,843,655, 5,814,454, 5,837,196, 5,436,327, 5,412,087, 5,405,783, the disclosures of which are incorporated herein by reference in their entireties.

In an alternative embodiment, a “microarray” may also refer to a “peptide microarray” or “protein microarray” having a substrate-bound collection of plurality of polypeptides, the binding to each of the plurality of bound polypeptides being separately detectable. Alternatively, the peptide microarray may have a plurality of binders, including but not limited to monoclonal antibodies, polyclonal antibodies, phage display binders, yeast 2 hybrid binders, aptamers, which can specifically detect the binding of the polypeptides of this invention. The array may be based on autoantibody detection to the polypeptides of this invention, see Robinson *et al.*, *Nature Medicine* 8(3):295-301 (2002). Examples of peptide arrays may be found in WO 02/31463, WO 02/25288, WO 01/94946, WO 01/88162, WO 01/68671, WO 01/57259, WO 00/61806, WO 00/54046, WO 00/47774, WO 99/40434, WO 99/39210, WO 97/42507 and U.S. Patent Nos. 6,268,210, 5,766,960, 5,143,854, the disclosures of which are incorporated herein by reference in their entireties.

In addition, determination of the levels of the CSNA or CSP may be made in a multiplex manner using techniques described in WO 02/29109, WO 02/24959, WO 01/83502, WO01/73113, WO 01/59432, WO 01/57269, WO 99/67641, the disclosures of which are incorporated herein by reference in their entireties.

5 The term “mutant”, “mutated”, or “mutation” when applied to nucleic acid sequences means that nucleotides in a nucleic acid sequence may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci
10 within a nucleic acid sequence. In a preferred embodiment of the present invention, the nucleic acid sequence is the wild type nucleic acid sequence encoding a CSP or is a CSNA. The nucleic acid sequence may be mutated by any method known in the art including those mutagenesis techniques described *infra*.

 The term “error-prone PCR” refers to a process for performing PCR under
15 conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. *See, e.g.,* Leung *et al.*, *Technique* 1: 11-15 (1989) and Caldwell *et al.*, *PCR Methods Applic.* 2: 28-33 (1992).

 The term “oligonucleotide-directed mutagenesis” refers to a process which enables
20 the generation of site-specific mutations in any cloned DNA segment of interest. *See, e.g.,* Reidhaar-Olson *et al.*, *Science* 241: 53-57 (1988).

 The term “assembly PCR” refers to a process which involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR reactions occur in parallel in the same vial, with the products of one reaction priming the
25 products of another reaction.

 The term “sexual PCR mutagenesis” or “DNA shuffling” refers to a method of error-prone PCR coupled with forced homologous recombination between DNA molecules of different but highly related DNA sequence *in vitro*, caused by random fragmentation of the DNA molecule based on sequence similarity, followed by fixation of
30 the crossover by primer extension in an error-prone PCR reaction. *See, e.g.,* Stemmer, *Proc. Natl. Acad. Sci. U.S.A.* 91: 10747-10751 (1994). DNA shuffling can be carried out between several related genes (“Family shuffling”).

The term “*in vivo* mutagenesis” refers to a process of generating random mutations in any cloned DNA of interest which involves the propagation of the DNA in a strain of bacteria such as *E. coli* that carries mutations in one or more of the DNA repair pathways. These “mutator” strains have a higher random mutation rate than that of a wild-type parent. Propagating the DNA in a mutator strain will eventually generate random mutations within the DNA.

The term “cassette mutagenesis” refers to any process for replacing a small region of a double-stranded DNA molecule with a synthetic oligonucleotide “cassette” that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence.

The term “recursive ensemble mutagenesis” refers to an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. See, e.g., Arkin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 89: 7811-7815 (1992).

The term “exponential ensemble mutagenesis” refers to a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. See, e.g., Delegrave *et al.*, *Biotechnology Research* 11: 1548-1552 (1993); Arnold, *Current Opinion in Biotechnology* 4: 450-455 (1993).

“Operatively linked” expression control sequences refers to a linkage in which the expression control sequence is either contiguous with the gene of interest to control the gene of interest, or acts in *trans* or at a distance to control the gene of interest.

The term “expression control sequence” as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination, promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (e.g., ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature

of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and transcription termination sequence. The term "control sequences" is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

The term "vector," as used herein, is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Viral vectors that infect bacterial cells are referred to as bacteriophages. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression vectors"). In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include other forms of expression vectors that serve equivalent functions.

The term "recombinant host cell" (or simply "host cell"), as used herein, is intended to refer to a cell into which a recombinant expression vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein.

As used herein, the phrase "open reading frame" and the equivalent acronym "ORF" refers to that portion of a transcript-derived nucleic acid that can be translated in its

entirety into a sequence of contiguous amino acids. As so defined, an ORF has length, measured in nucleotides, exactly divisible by 3. As so defined, an ORF need not encode the entirety of a natural protein.

As used herein, the phrase "ORF-encoded peptide" refers to the predicted or actual
5 translation of an ORF.

As used herein, the phrase "degenerate variant" of a reference nucleic acid sequence is meant to be inclusive of all nucleic acid sequences that can be directly translated, using the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence.

10 The term "polypeptide" encompasses both naturally-occurring and non-naturally-occurring proteins and polypeptides, as well as polypeptide fragments and polypeptide mutants, derivatives and analogs thereof. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different modules within a single polypeptide each of which has one or more distinct activities. A preferred polypeptide in
15 accordance with the invention comprises a CSP encoded by a nucleic acid molecule of the instant invention, or a fragment, mutant, analog and derivative thereof.

The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide that by virtue of its origin or source of derivation (1) is not associated with naturally associated components that accompany it in its native state, (2) is free of other proteins
20 from the same species (3) is expressed by a cell from a different species, or (4) does not occur in nature. Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be "isolated" from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally associated components by isolation, using protein
25 purification techniques well-known in the art.

A protein or polypeptide is "substantially pure," "substantially homogeneous" or "substantially purified" when at least about 60% to 75% of a sample exhibits a single species of polypeptide. The polypeptide or protein may be monomeric or multimeric. A substantially pure polypeptide or protein will typically comprise about 50%, 60%, 70%,
30 80% or 90% W/W of a protein sample, more usually about 95%, and preferably will be over 99% pure. Protein purity or homogeneity may be determined by a number of means well-known in the art, such as polyacrylamide gel electrophoresis of a protein sample, followed by visualizing a single polypeptide band upon staining the gel with a stain well-

known in the art. For certain purposes, higher resolution may be provided by using HPLC or other means well-known in the art for purification.

The term "fragment" when used herein with respect to polypeptides of the present invention refers to a polypeptide that has an amino-terminal and/or
5 carboxy-terminal deletion compared to a full-length CSP. In a preferred embodiment, the fragment is a contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally-occurring polypeptide. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or
10 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

A "derivative" when used herein with respect to polypeptides of the present invention refers to a polypeptide which is substantially similar in primary structural sequence to a CSP but which include, *e.g.*, *in vivo* or *in vitro* chemical and biochemical
15 modifications that are not found in the CSP. Such modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation,
20 formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.
25 Other modification include, *e.g.*, labeling with radionuclides, and various enzymatic modifications, as will be readily appreciated by those skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well-known in the art, and include radioactive isotopes such as ¹²⁵I, ³²P, ³⁵S, ¹⁴C and ³H, ligands which bind to labeled antiligands (*e.g.*, antibodies), fluorophores,
30 chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation.

Methods for labeling polypeptides are well-known in the art. *See* Ausubel (1992), *supra*; Ausubel (1999), *supra*.

The term “fusion protein” refers to polypeptides of the present invention coupled to a heterologous amino acid sequences. Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusion proteins can be produced recombinantly by constructing a nucleic acid sequence that encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the polypeptide or a fragment thereof to another protein.

The term “analog” refers to both polypeptide analogs and non-peptide analogs. The term “polypeptide analog” as used herein refers to a polypeptide that is comprised of a segment of at least 25 amino acids that has substantial identity to a portion of an amino acid sequence but which contains non-natural amino acids or non-natural inter-residue bonds. In a preferred embodiment, the analog has the same or similar biological activity as the native polypeptide. Typically, polypeptide analogs comprise a conservative amino acid substitution (or insertion or deletion) with respect to the naturally-occurring sequence. Analog typically are at least 20 amino acids long, preferably at least 50 amino acids long or longer, and can often be as long as a full-length naturally-occurring polypeptide.

The term “non-peptide analog” refers to a compound with properties that are analogous to those of a reference polypeptide. A non-peptide compound may also be termed a “peptide mimetic” or a “peptidomimetic.” Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to useful peptides may be used to produce an equivalent effect. Generally, peptidomimetics are structurally similar to a paradigm polypeptide (*i.e.*, a polypeptide that has a desired biochemical property or pharmacological activity), but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of: --CH₂NH--, --CH₂S--, --CH₂-CH₂--, --CH=CH--(cis and trans), --COCH₂--, --CH(OH)CH₂--, and --CH₂SO--, by methods well-known in the art. Systematic substitution of one or more amino acids of a consensus sequence with a D-amino acid of

the same type (*e.g.*, D-lysine in place of L-lysine) may also be used to generate more stable peptides. In addition, constrained peptides comprising a consensus sequence or a substantially identical consensus sequence variation may be generated by methods known in the art (Rizo *et al.*, *Ann. Rev. Biochem.* 61:387-418 (1992)). For example, one may add
5 internal cysteine residues capable of forming intramolecular disulfide bridges which cyclize the peptide.

The term “mutant” or “mutein” when referring to a polypeptide of the present invention relates to an amino acid sequence containing substitutions, insertions or deletions of one or more amino acids compared to the amino acid sequence of a CSP. A
10 mutein may have one or more amino acid point substitutions, in which a single amino acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally-occurring protein, and/or truncations of the amino acid sequence at either or both the amino or carboxy termini. Further, a mutein may have the same or
15 different biological activity as the naturally-occurring protein. For instance, a mutein may have an increased or decreased biological activity. A mutein has at least 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are muteins having 80%, 85% or 90% sequence similarity to a CSP. In an even more preferred embodiment, a mutein exhibits
20 95% sequence identity, even more preferably 97%, even more preferably 98% and even more preferably 99%. Sequence similarity may be measured by any common sequence analysis algorithm, such as Gap or Bestfit.

Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming
25 protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such analogs. For example, single or multiple amino acid substitutions (preferably conservative amino acid substitutions) may be made in the naturally-occurring sequence (preferably in the portion of the polypeptide outside the domain(s) forming intermolecular contacts. In a preferred
30 embodiment, the amino acid substitutions are moderately conservative substitutions or conservative substitutions. In a more preferred embodiment, the amino acid substitutions are conservative substitutions. A conservative amino acid substitution should not substantially change the structural characteristics of the parent sequence (*e.g.*, a

replacement amino acid should not tend to disrupt a helix that occurs in the parent sequence, or disrupt other types of secondary structure that characterizes the parent sequence). Examples of art-recognized polypeptide secondary and tertiary structures are described in Creighton (ed.), Proteins, Structures and Molecular Principles, W. H.

- 5 Freeman and Company (1984); Branden *et al.* (ed.), Introduction to Protein Structure, Garland Publishing (1991); Thornton *et al.*, *Nature* 354:105-106 (1991).

As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. See Golub *et al.* (eds.), Immunology - A Synthesis 2nd Ed., Sinauer Associates (1991). Stereoisomers (*e.g.*, D-amino acids) of the twenty conventional amino
10 acids, unnatural amino acids such as α -, α -disubstituted amino acids, N-alkyl amino acids, and other unconventional amino acids may also be suitable components for polypeptides of the present invention. Examples of unconventional amino acids include:
4-hydroxyproline, γ -carboxyglutamate, ϵ -N,N,N-trimethyllysine, ϵ -N-acetyllysine,
O-phosphoserine, N-acetylserine, N-formylmethionine, 3-methylhistidine,
15 5-hydroxylysine, s-N-methylarginine, and other similar amino acids and imino acids (*e.g.*, 4-hydroxyproline). In the polypeptide notation used herein, the lefthand direction is the amino terminal direction and the right hand direction is the carboxy-terminal direction, in accordance with standard usage and convention.

By "homology" or "homologous" when referring to a polypeptide of the present
20 invention it is meant polypeptides from different organisms with a similar sequence to the encoded amino acid sequence of a CSP and a similar biological activity or function. Although two polypeptides are said to be "homologous," this does not imply that there is necessarily an evolutionary relationship between the polypeptides. Instead, the term "homologous" is defined to mean that the two polypeptides have similar amino acid
25 sequences and similar biological activities or functions. In a preferred embodiment, a homologous polypeptide is one that exhibits 50% sequence similarity to CSP, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are homologous polypeptides that exhibit 80%, 85% or 90% sequence similarity to a CSP. In a yet more preferred embodiment, a homologous polypeptide exhibits 95%, 97%, 98%
30 or 99% sequence similarity.

When "sequence similarity" is used in reference to polypeptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions. In a preferred embodiment, a polypeptide that has "sequence similarity"

comprises conservative or moderately conservative amino acid substitutions. A “conservative amino acid substitution” is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (*e.g.*, charge or hydrophobicity). In general, a conservative amino acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of similarity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well-known to those of skill in the art. *See, e.g.*, Pearson, *Methods Mol. Biol.* 24: 307-31 (1994).

For instance, the following six groups each contain amino acids that are conservative substitutions for one another:

- 1) Serine (S), Threonine (T);
- 2) Aspartic Acid (D), Glutamic Acid (E);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

Alternatively, a conservative replacement is any change having a positive value in the PAM250 log-likelihood matrix disclosed in Gonnet *et al.*, *Science* 256: 1443-45 (1992). A “moderately conservative” replacement is any change having a nonnegative value in the PAM250 log-likelihood matrix.

Sequence similarity for polypeptides, which is also referred to as sequence identity, is typically measured using sequence analysis software. Protein analysis software matches similar sequences using measures of similarity assigned to various substitutions, deletions and other modifications, including conservative amino acid substitutions. For instance, GCG contains programs such as “Gap” and “Bestfit” which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or between a wild type protein and a mutein thereof. *See, e.g.*, GCG Version 6.1. Other programs include FASTA, discussed *supra*.

A preferred algorithm when comparing a sequence of the invention to a database containing a large number of sequences from different organisms is the computer program

BLAST, especially blastp or tblastn. *See, e.g., Altschul et al., J. Mol. Biol.* 215: 403-410 (1990); Altschul *et al., Nucleic Acids Res.* 25:3389-402 (1997). Preferred parameters for blastp are:

| | | |
|----|-----------------------|---------------|
| | Expectation value: | 10 (default) |
| 5 | Filter: | seg (default) |
| | Cost to open a gap: | 11 (default) |
| | Cost to extend a gap: | 1 (default) |
| | Max. alignments: | 100 (default) |
| | Word size: | 11 (default) |
| 10 | No. of descriptions: | 100 (default) |
| | Penalty Matrix: | BLOSUM62 |

The length of polypeptide sequences compared for homology will generally be at least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number of different organisms, it is preferable to compare amino acid sequences.

Algorithms other than blastp for database searching using amino acid sequences are known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA (*e.g.*, FASTA2 and FASTA3) provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson (1990), *supra*; Pearson (2000), *supra*. For example, percent sequence identity between amino acid sequences can be determined using FASTA with its default or recommended parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1.

An "antibody" refers to an intact immunoglobulin, or to an antigen-binding portion thereof that competes with the intact antibody for specific binding to a molecular species, *e.g.*, a polypeptide of the instant invention. Antigen-binding portions may be produced by recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies. Antigen-binding portions include, *inter alia*, Fab, Fab', F(ab')₂, Fv, dAb, and complementarity determining region (CDR) fragments, single-chain antibodies (scFv), chimeric antibodies, diabodies and polypeptides that contain at least a portion of an immunoglobulin that is sufficient to confer specific antigen binding to the polypeptide. A Fab fragment is a monovalent fragment consisting of the VL, VH, CL and CH1 domains;

a F(ab')₂ fragment is a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; a Fd fragment consists of the VH and CH1 domains; a Fv fragment consists of the VL and VH domains of a single arm of an antibody; and a dAb fragment consists of a VH domain. *See, e.g., Ward et al., Nature* 341: 544-546 (1989).

5 By "bind specifically" and "specific binding" as used herein it is meant the ability of the antibody to bind to a first molecular species in preference to binding to other molecular species with which the antibody and first molecular species are admixed. An antibody is said specifically to "recognize" a first molecular species when it can bind specifically to that first molecular species.

10 A single-chain antibody (scFv) is an antibody in which VL and VH regions are paired to form a monovalent molecule via a synthetic linker that enables them to be made as a single protein chain. *See, e.g., Bird et al., Science* 242: 423-426 (1988); Huston et al., *Proc. Natl. Acad. Sci. USA* 85: 5879-5883 (1988). Diabodies are bivalent, bispecific antibodies in which VH and VL domains are expressed on a single polypeptide chain, but
15 using a linker that is too short to allow for pairing between the two domains on the same chain, thereby forcing the domains to pair with complementary domains of another chain and creating two antigen binding sites. *See e.g., Holliger et al., Proc. Natl. Acad. Sci. USA* 90: 6444-6448 (1993); Poljak et al., *Structure* 2: 1121-1123 (1994). One or more CDRs may be incorporated into a molecule either covalently or noncovalently to make it an
20 immunoadhesin. An immunoadhesin may incorporate the CDR(s) as part of a larger polypeptide chain, may covalently link the CDR(s) to another polypeptide chain, or may incorporate the CDR(s) noncovalently. The CDRs permit the immunoadhesin to specifically bind to a particular antigen of interest. A chimeric antibody is an antibody that contains one or more regions from one antibody and one or more regions from one or
25 more other antibodies.

An antibody may have one or more binding sites. If there is more than one binding site, the binding sites may be identical to one another or may be different. For instance, a naturally-occurring immunoglobulin has two identical binding sites, a single-chain antibody or Fab fragment has one binding site, while a "bispecific" or "bifunctional"
30 antibody has two different binding sites.

An "isolated antibody" is an antibody that (1) is not associated with naturally-associated components, including other naturally-associated antibodies, that accompany it in its native state, (2) is free of other proteins from the same species, (3) is expressed by a

cell from a different species, or (4) does not occur in nature. It is known that purified proteins, including purified antibodies, may be stabilized with non-naturally-associated components. The non-naturally-associated component may be a protein, such as albumin (*e.g.*, BSA) or a chemical such as polyethylene glycol (PEG).

5 A “neutralizing antibody” or “an inhibitory antibody” is an antibody that inhibits the activity of a polypeptide or blocks the binding of a polypeptide to a ligand that normally binds to it. An “activating antibody” is an antibody that increases the activity of a polypeptide.

10 The term “epitope” includes any protein determinant capable of specific binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three dimensional structural characteristics, as well as specific charge characteristics. An antibody is said to specifically bind an antigen when the dissociation constant is less than 1 μ M, preferably less than 100 nM and most preferably
15 less than 10 nM.

 The term “patient” includes human and veterinary subjects.

 Throughout this specification and claims, the word “comprise,” or variations such as “comprises” or “comprising,” will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

20 The term “colon specific” refers to a nucleic acid molecule or polypeptide that is expressed predominantly in the colon as compared to other tissues in the body. In a preferred embodiment, a “colon specific” nucleic acid molecule or polypeptide is detected at a level that is 2-fold higher than any other tissue in the body. In a more preferred embodiment, the “colon specific” nucleic acid molecule or polypeptide is detected at a
25 level that is 5-fold higher than any other tissue in the body, more preferably at least 10-fold, 15-fold, 20-fold, 25-fold, 50-fold or 100-fold higher than any other tissue in the body. Nucleic acid molecule levels may be measured by nucleic acid hybridization, such as Northern blot hybridization, or quantitative PCR. Polypeptide levels may be measured by any method known to accurately quantitate protein levels, such as Western blot
30 analysis.

Nucleic Acid Molecules, Regulatory Sequences, Vectors, Host Cells and Recombinant Methods of Making Polypeptides

Nucleic Acid Molecules

5 One aspect of the invention provides isolated nucleic acid molecules that are specific to the colon or to colon cells or tissue or that are derived from such nucleic acid molecules. These isolated colon specific nucleic acids (CSNAs) may comprise cDNA genomic DNA, RNA, or a combination thereof, a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. A CSNA may be derived
10 from an animal. In a preferred embodiment, the CSNA is derived from a human or other mammal. In a more preferred embodiment, the CSNA is derived from a human or other primate. In an even more preferred embodiment, the CSNA is derived from a human.

 In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to colon, a colon-specific polypeptide (CSP). In a more preferred embodiment,
15 the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 101-190. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1-100. Nucleotide sequences of the instantly-described nucleic acid molecules were determined by sequencing a DNA molecule that had resulted, directly or indirectly, from at least one enzymatic
20 polymerization reaction (*e.g.*, reverse transcription and/or polymerase chain reaction) using an automated sequencer (such as the MegaBACE™ 1000, Amersham Biosciences, Sunnyvale, CA, USA).

 Nucleic acid molecules of the present invention may also comprise sequences that selectively hybridizes to a nucleic acid molecule encoding a CSNA or a complement or
25 antisense thereof. The hybridizing nucleic acid molecule may or may not encode a polypeptide or may or may not encode a CSP. However, in a preferred embodiment, the hybridizing nucleic acid molecule encodes a CSP. In a more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule that encodes a polypeptide
30 comprising an amino acid sequence of SEQ ID NO: 101-190. In an even more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule comprising the nucleic acid sequence of SEQ ID NO: 1-100 or the antisense sequence thereof. Preferably, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule encoding a

CSP under low stringency conditions. More preferably, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule encoding a CSP under moderate stringency conditions. Most preferably, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule encoding a CSP under high stringency conditions. In a preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of SEQ ID NO: 101-190. In a more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule or the antisense sequence of a nucleic acid molecule comprising a nucleic acid sequence selected from SEQ ID NO: 1-100.

Nucleic acid molecules of the present invention may also comprise nucleic acid sequences that exhibit substantial sequence similarity to a nucleic acid encoding a CSP or a complement of the encoding nucleic acid molecule. In this embodiment, it is preferred that the nucleic acid molecule exhibit substantial sequence similarity to a nucleic acid molecule encoding human CSP. More preferred is a nucleic acid molecule exhibiting substantial sequence similarity to a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 101-190. By substantial sequence similarity it is meant a nucleic acid molecule having at least 60% sequence identity with a nucleic acid molecule encoding a CSP, such as a polypeptide having an amino acid sequence of SEQ ID NO: 101-190, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the similar nucleic acid molecule is one that has at least 90% sequence identity with a nucleic acid molecule encoding a CSP, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. Most preferred in this embodiment is a nucleic acid molecule that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a nucleic acid molecule encoding a CSP.

The nucleic acid molecules of the present invention are also inclusive of those exhibiting substantial sequence similarity to a CSNA or its complement. In this embodiment, it is preferred that the nucleic acid molecule exhibit substantial sequence similarity to a nucleic acid molecule having a nucleic acid sequence of SEQ ID NO: 1-100. By substantial sequence similarity it is meant a nucleic acid molecule that has at

least 60% sequence identity with a CSNA, such as one having a nucleic acid sequence of SEQ ID NO: 1-100, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. More preferred is a nucleic acid molecule that has at least 90% sequence identity with a CSNA, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. Most preferred is a nucleic acid molecule that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a CSNA.

Nucleic acid molecules that exhibit substantial sequence similarity are inclusive of sequences that exhibits sequence identity over their entire length to a CSNA or to a nucleic acid molecule encoding a CSP, as well as sequences that are similar over only a part of its length. In this case, the part is at least 50 nucleotides of the CSNA or the nucleic acid molecule encoding a CSP, preferably at least 100 nucleotides, more preferably at least 150 or 200 nucleotides, even more preferably at least 250 or 300 nucleotides, still more preferably at least 400 or 500 nucleotides.

The substantially similar nucleic acid molecule may be a naturally-occurring one that is derived from another species, especially one derived from another primate, wherein the similar nucleic acid molecule encodes an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 101-190 or demonstrates significant sequence identity to the nucleotide sequence of SEQ ID NO: 1-100. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule from a human, when the CSNA is a member of a gene family. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, hamster, cow, horse and pig; and wild animals, *e.g.*, monkey, fox, lions, tigers, bears, giraffes, zebras, etc. The substantially similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring substantially similar nucleic acid molecule may be isolated directly from humans or other species. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by random mutation of a nucleic acid molecule. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by directed mutation of a CSNA. In a preferred embodiment, the substantially similar nucleic acid molecule is an CSNA.

The nucleic acid molecules of the present invention are also inclusive of allelic variants of a CSNA or a nucleic acid encoding a CSP. For example, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes and the sequence determined from one individual of a species may differ from other allelic forms present within the population. More than 1.4 million SNPs have already identified in the human genome, International Human Genome Sequencing Consortium, *Nature* 409: 860-921 (2001) – Variants with small deletions and insertions of more than a single nucleotide are also found in the general population, and often do not alter the function of the protein. In addition, amino acid substitutions occur frequently among natural allelic variants, and often do not substantially change protein function.

In a preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that encodes a CSP. In a more preferred embodiment, the gene is transcribed into an mRNA that encodes a CSP comprising an amino acid sequence of SEQ ID NO: 101-190. In another preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that is a CSNA. In a more preferred embodiment, the gene is transcribed into an mRNA that comprises the nucleic acid sequence of SEQ ID NO: 1-100. Also preferred is that the allelic variant is a naturally-occurring allelic variant in the species of interest, particularly human.

Nucleic acid molecules of the present invention are also inclusive of nucleic acid sequences comprising a part of a nucleic acid sequence of the instant invention. The part may or may not encode a polypeptide, and may or may not encode a polypeptide that is a CSP. In a preferred embodiment, the part encodes a CSP. In one embodiment, the nucleic acid molecule comprises a part of a CSNA. In another embodiment, the nucleic acid molecule comprises a part of a nucleic acid molecule that hybridizes or exhibits substantial sequence similarity to a CSNA. In another embodiment, the nucleic acid molecule comprises a part of a nucleic acid molecule that is an allelic variant of a CSNA. In yet another embodiment, the nucleic acid molecule comprises a part of a nucleic acid molecule that encodes a CSP. A part comprises at least 10 nucleotides, more preferably at least 15, 17, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides. The maximum size of a nucleic acid part is one nucleotide shorter than the sequence of the nucleic acid molecule encoding the full-length protein.

Nucleic acid molecules of the present invention are also inclusive of nucleic acid sequences that encode fusion proteins, homologous proteins, polypeptide fragments, muteins and polypeptide analogs, as described *infra*.

Nucleic acid molecules of the present invention are also inclusive of nucleic acid sequences containing modifications of the native nucleic acid molecule. Examples of such modifications include, but are not limited to, nonnative internucleoside bonds, post-synthetic modifications or altered nucleotide analogues. One having ordinary skill in the art would recognize that the type of modification that can be made will depend upon the intended use of the nucleic acid molecule. For instance, when the nucleic acid molecule is used as a hybridization probe, the range of such modifications will be limited to those that permit sequence-discriminating base pairing of the resulting nucleic acid. When used to direct expression of RNA or protein *in vitro* or *in vivo*, the range of such modifications will be limited to those that permit the nucleic acid to function properly as a polymerization substrate. When the isolated nucleic acid is used as a therapeutic agent, the modifications will be limited to those that do not confer toxicity upon the isolated nucleic acid.

Accordingly, in one embodiment, a nucleic acid molecule may include nucleotide analogues that incorporate labels that are directly detectable, such as radiolabels or fluorophores, or nucleotide analogues that incorporate labels that can be visualized in a subsequent reaction, such as biotin or various haptens. The labeled nucleic acid molecules are particularly useful as hybridization probes.

Common radiolabeled analogues include those labeled with ^{33}P , ^{32}P , and ^{35}S , such as α - ^{32}P -dATP, α - ^{32}P -dCTP, α - ^{32}P -dGTP, α - ^{32}P -dTTP, α - ^{32}P -3'-dATP, α - ^{32}P -ATP, α - ^{32}P -CTP, α - ^{32}P -GTP, α - ^{32}P -UTP, α - ^{35}S -dATP, γ - ^{35}S -GTP, γ - ^{33}P -dATP, and the like.

Commercially available fluorescent nucleotide analogues readily incorporated into the nucleic acids of the present invention include Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy3-dUTP (Amersham Biosciences, Piscataway, New Jersey, USA), fluorescein-12-dUTP, tetramethylrhodamine-6-dUTP, Texas Red®-5-dUTP, Cascade Blue®-7-dUTP, BODIPY® FL-14-dUTP, BODIPY® TMR-14-dUTP, BODIPY® TR-14-dUTP, Rhodamine Green™-5-dUTP, Oregon Green® 488-5-dUTP, Texas Red®-12-dUTP, BODIPY® 630/650-14-dUTP, BODIPY® 650/665-14-dUTP, Alexa Fluor® 488-5-dUTP, Alexa Fluor® 532-5-dUTP, Alexa Fluor® 568-5-dUTP, Alexa Fluor® 594-5-dUTP, Alexa Fluor® 546-14-dUTP, fluorescein-12-UTP, tetramethylrhodamine-6-UTP, Texas

Red®-5-UTP, Cascade Blue®-7-UTP, BODIPY® FL-14-UTP, BODIPY® TMR-14-UTP, BODIPY® TR-14-UTP, Rhodamine Green™-5-UTP, Alexa Fluor® 488-5-UTP, Alexa Fluor® 546-14-UTP (Molecular Probes, Inc. Eugene, OR, USA). One may also custom synthesize nucleotides having other fluorophores. See Henegariu *et al.*, *Nature*

5 *Biotechnol.* 18: 345-348 (2000).

Haptens that are commonly conjugated to nucleotides for subsequent labeling include biotin (biotin-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA; biotin-21-UTP, biotin-21-dUTP, Clontech Laboratories, Inc., Palo Alto, CA, USA), digoxigenin (DIG-11-dUTP, alkali labile, DIG-11-UTP, Roche Diagnostics Corp., Indianapolis, IN, USA), and dinitrophenyl (dinitrophenyl-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA).

Nucleic acid molecules of the present invention can be labeled by incorporation of labeled nucleotide analogues into the nucleic acid. Such analogues can be incorporated by enzymatic polymerization, such as by nick translation, random priming, polymerase chain reaction (PCR), terminal transferase tailing, and end-filling of overhangs, for DNA molecules, and *in vitro* transcription driven, *e.g.*, from phage promoters, such as T7, T3, and SP6, for RNA molecules. Commercial kits are readily available for each such labeling approach. Analogues can also be incorporated during automated solid phase chemical synthesis. Labels can also be incorporated after nucleic acid synthesis, with the 5' phosphate and 3' hydroxyl providing convenient sites for post-synthetic covalent attachment of detectable labels.

Other post-synthetic approaches also permit internal labeling of nucleic acids. For example, fluorophores can be attached using a cisplatin reagent that reacts with the N7 of guanine residues (and, to a lesser extent, adenine bases) in DNA, RNA, and Peptide Nucleic Acids (PNA) to provide a stable coordination complex between the nucleic acid and fluorophore label (Universal Linkage System) (available from Molecular Probes, Inc., Eugene, OR, USA and Amersham Pharmacia Biotech, Piscataway, NJ, USA); see Alers *et al.*, *Genes, Chromosomes & Cancer* 25: 301-305 (1999); Jelsma *et al.*, *J. NIH Res.* 5: 82 (1994); Van Belkum *et al.*, *BioTechniques* 16: 148-153 (1994). Alternatively, nucleic acids can be labeled using a disulfide-containing linker (FastTag™ Reagent, Vector Laboratories, Inc., Burlingame, CA, USA) that is photo- or thermally coupled to the target nucleic acid using aryl azide chemistry; after reduction, a free thiol is available for coupling to a hapten, fluorophore, sugar, affinity ligand, or other marker.

One or more independent or interacting labels can be incorporated into the nucleic acid molecules of the present invention. For example, both a fluorophore and a moiety that in proximity thereto acts to quench fluorescence can be included to report specific hybridization through release of fluorescence quenching or to report exonucleotidic excision. See, e.g., Tyagi *et al.*, *Nature Biotechnol.* 14: 303-308 (1996); Tyagi *et al.*, *Nature Biotechnol.* 16: 49-53 (1998); Sokol *et al.*, *Proc. Natl. Acad. Sci. USA* 95: 11538-11543 (1998); Kostrikis *et al.*, *Science* 279: 1228-1229 (1998); Marras *et al.*, *Genet. Anal.* 14: 151-156 (1999); Holland *et al.*, *Proc. Natl. Acad. Sci. USA* 88: 7276-7280 (1991); Heid *et al.*, *Genome Res.* 6(10): 986-94 (1996); Kuimelis *et al.*, *Nucleic Acids Symp. Ser.* (37): 255-6 (1997); and United States Patent Nos. 5,846,726, 5,925,517, 5,925,517, 5,723,591 and 5,538,848, the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules of the present invention may also be modified by altering one or more native phosphodiester internucleoside bonds to more nuclease-resistant, internucleoside bonds. See Hartmann *et al.* (eds.), Manual of Antisense Methodology: Perspectives in Antisense Science, Kluwer Law International (1999); Stein *et al.* (eds.), Applied Antisense Oligonucleotide Technology, Wiley-Liss (1998); Chadwick *et al.* (eds.), Oligonucleotides as Therapeutic Agents - Symposium No. 209, John Wiley & Son Ltd (1997). Such altered internucleoside bonds are often desired for antisense techniques or for targeted gene correction, Gamper *et al.*, *Nucl. Acids Res.* 28(21): 4332-4339 (2000).

Modified oligonucleotide backbones include, without limitation, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Representative United States patents that teach the preparation of the above phosphorus-containing linkages include, but are not limited to, United States Patent Nos. 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361;

and 5,625,050, the disclosures of which are incorporated herein by reference in their entireties. In a preferred embodiment, the modified internucleoside linkages may be used for antisense techniques.

Other modified oligonucleotide backbones do not include a phosphorus atom, but
5 have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and
10 thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH₂ component parts. Representative U.S. patents that teach the preparation of the above backbones include, but are not limited to, United States Patent
15 Nos. 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437 and 5,677,439; the disclosures of which are incorporated herein by reference in their entireties.

20 In other preferred nucleic acid molecules, both the sugar and the internucleoside linkage are replaced with novel groups, such as peptide nucleic acids (PNA). In PNA compounds, the phosphodiester backbone of the nucleic acid is replaced with an amide-containing backbone, in particular by repeating N-(2-aminoethyl) glycine units linked by amide bonds. Nucleobases are bound directly or indirectly to aza nitrogen atoms of the
25 amide portion of the backbone, typically by methylene carbonyl linkages. PNA can be synthesized using a modified peptide synthesis protocol. PNA oligomers can be synthesized by both Fmoc and tBoc methods. Representative U.S. patents that teach the preparation of PNA compounds include, but are not limited to, United States Patent Nos. 5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference in
30 its entirety. Automated PNA synthesis is readily achievable on commercial synthesizers (*see, e.g.*, "PNA User's Guide," Rev. 2, February 1998, Perseptive Biosystems Part No. 60138, Applied Biosystems, Inc., Foster City, CA). PNA molecules are advantageous for a number of reasons. First, because the PNA backbone is uncharged, PNA/DNA and

PNA/RNA duplexes have a higher thermal stability than is found in DNA/DNA and DNA/RNA duplexes. The T_m of a PNA/DNA or PNA/RNA duplex is generally 1°C higher per base pair than the T_m of the corresponding DNA/DNA or DNA/RNA duplex (in 100 mM NaCl). Second, PNA molecules can also form stable PNA/DNA complexes at low ionic strength, under conditions in which DNA/DNA duplex formation does not occur. Third, PNA also demonstrates greater specificity in binding to complementary DNA because a PNA/DNA mismatch is more destabilizing than DNA/DNA mismatch. A single mismatch in mixed a PNA/DNA 15-mer lowers the T_m by 8–20°C (15°C on average). In the corresponding DNA/DNA duplexes, a single mismatch lowers the T_m by 4–16°C (11°C on average). Because PNA probes can be significantly shorter than DNA probes, their specificity is greater. Fourth, PNA oligomers are resistant to degradation by enzymes, and the lifetime of these compounds is extended both *in vivo* and *in vitro* because nucleases and proteases do not recognize the PNA polyamide backbone with nucleobase sidechains. See, e.g., Ray *et al.*, *FASEB J.* 14(9): 1041-60 (2000); Nielsen *et al.*, *Pharmacol Toxicol.* 86(1): 3-7 (2000); Larsen *et al.*, *Biochim Biophys Acta.* 1489(1): 159-66 (1999); Nielsen, *Curr. Opin. Struct. Biol.* 9(3): 353-7 (1999), and Nielsen, *Curr. Opin. Biotechnol.* 10(1): 71-5 (1999).

Nucleic acid molecules may be modified compared to their native structure throughout the length of the nucleic acid molecule or can be localized to discrete portions thereof. As an example of the latter, chimeric nucleic acids can be synthesized that have discrete DNA and RNA domains and that can be used for targeted gene repair and modified PCR reactions, as further described in, Misra *et al.*, *Biochem.* 37: 1917-1925 (1998); and Finn *et al.*, *Nucl. Acids Res.* 24: 3357-3363 (1996), and United States Patent Nos. 5,760,012 and 5,731,181, the disclosures of which are incorporated herein by reference in their entireties.

Unless otherwise specified, nucleic acid molecules of the present invention can include any topological conformation appropriate to the desired use; the term thus explicitly comprehends, among others, single-stranded, double-stranded, triplexed, quadruplexed, partially double-stranded, partially-triplexed, partially-quadruplexed, branched, hairpinned, circular, and padlocked conformations. Padlock conformations and their utilities are further described in Banér *et al.*, *Curr. Opin. Biotechnol.* 12: 11-15 (2001); Escude *et al.*, *Proc. Natl. Acad. Sci. USA* 14: 96(19):10603-7 (1999); and Nilsson *et al.*, *Science* 265(5181): 2085-8 (1994). Triplex and quadruplex conformations, and

their utilities, are reviewed in Praseuth *et al.*, *Biochim. Biophys. Acta.* 1489(1): 181-206 (1999); Fox, *Curr. Med. Chem.* 7(1): 17-37 (2000); Kochetkova *et al.*, *Methods Mol. Biol.* 130: 189-201 (2000); Chan *et al.*, *J. Mol. Med.* 75(4): 267-82 (1997).

5 *Methods for Using Nucleic Acid Molecules as Probes and Primers*

The isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize, and quantify hybridizing nucleic acids in, and isolate hybridizing nucleic acids from, both genomic and transcript-derived nucleic acid samples. When free in solution, such probes are typically, but not invariably, detectably
10 labeled; bound to a substrate, as in a microarray, such probes are typically, but not invariably unlabeled.

In one embodiment, the isolated nucleic acid molecules of the present invention can be used as probes to detect and characterize gross alterations in the gene of a CSNA, such as deletions, insertions, translocations, and duplications of the CSNA genomic locus
15 through fluorescence *in situ* hybridization (FISH) to chromosome spreads. *See, e.g.*, Andreeff *et al.* (eds.), Introduction to Fluorescence *In Situ* Hybridization: Principles and Clinical Applications, John Wiley & Sons (1999). The isolated nucleic acid molecules of the present invention can be used as probes to assess smaller genomic alterations using, *e.g.*, Southern blot detection of restriction fragment length polymorphisms. The isolated
20 nucleic acid molecules of the present invention can be used as probes to isolate genomic clones that include a nucleic acid molecule of the present invention, which thereafter can be restriction mapped and sequenced to identify deletions, insertions, translocations, and substitutions (single nucleotide polymorphisms, SNPs) at the sequence level.

The isolated nucleic acid molecules of the present invention can be also be used as
25 probes to detect, characterize, and quantify CSNA in, and isolate CSNA from, transcript-derived nucleic acid samples. In one embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by length, and quantify mRNA by Northern blot of total or poly-A⁺- selected RNA samples. In another embodiment, the isolated nucleic acid molecules of the present invention can be
30 used as hybridization probes to detect, characterize by location, and quantify mRNA by *in situ* hybridization to tissue sections. *See, e.g.*, Schwarchzacher *et al.*, In Situ Hybridization, Springer-Verlag New York (2000). In another preferred embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization

probes to measure the representation of clones in a cDNA library or to isolate hybridizing nucleic acid molecules acids from cDNA libraries, permitting sequence level characterization of mRNAs that hybridize to CSNAs, including, without limitations, identification of deletions, insertions, substitutions, truncations, alternatively spliced forms and single nucleotide polymorphisms. In yet another preferred embodiment, the nucleic acid molecules of the instant invention may be used in microarrays.

All of the aforementioned probe techniques are well within the skill in the art, and are described at greater length in standard texts such as Sambrook (2001), *supra*; Ausubel (1999), *supra*; and Walker *et al.* (eds.), The Nucleic Acids Protocols Handbook, Humana Press (2000).

In another embodiment, a nucleic acid molecule of the invention may be used as a probe or primer to identify and/or amplify a second nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of the invention. In this embodiment, it is preferred that the probe or primer be derived from a nucleic acid molecule encoding a CSP. More preferably, the probe or primer is derived from a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 101-190. Also preferred are probes or primers derived from a CSNA. More preferred are probes or primers derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-100.

In general, a probe or primer is at least 10 nucleotides in length, more preferably at least 12, more preferably at least 14 and even more preferably at least 16 or 17 nucleotides in length. In an even more preferred embodiment, the probe or primer is at least 18 nucleotides in length, even more preferably at least 20 nucleotides and even more preferably at least 22 nucleotides in length. Primers and probes may also be longer in length. For instance, a probe or primer may be 25 nucleotides in length, or may be 30, 40 or 50 nucleotides in length. Methods of performing nucleic acid hybridization using oligonucleotide probes are well-known in the art. *See, e.g.*, Sambrook *et al.*, 1989, *supra*, Chapter 11 and pp. 11.31-11.32 and 11.40-11.44, which describes radiolabeling of short probes, and pp. 11.45-11.53, which describe hybridization conditions for oligonucleotide probes, including specific conditions for probe hybridization (pp. 11.50-11.51).

Methods of performing primer-directed amplification are also well-known in the art. Methods for performing the polymerase chain reaction (PCR) are compiled, *inter alia*, in McPherson, PCR Basics: From Background to Bench, Springer Verlag (2000); Innis *et*

- al.* (eds.), PCR Applications: Protocols for Functional Genomics, Academic Press (1999); Gelfand *et al.* (eds.), PCR Strategies, Academic Press (1998); Newton *et al.*, PCR, Springer-Verlag New York (1997); Burke (ed.), PCR: Essential Techniques, John Wiley & Son Ltd (1996); White (ed.), PCR Cloning Protocols: From Molecular Cloning to
5 Genetic Engineering, Vol. 67, Humana Press (1996); and McPherson *et al.* (eds.), PCR 2: A Practical Approach, Oxford University Press, Inc. (1995). Methods for performing RT-PCR are collected, *e.g.*, in Siebert *et al.* (eds.), Gene Cloning and Analysis by RT-PCR, Eaton Publishing Company/Bio Techniques Books Division, 1998; and Siebert (ed.), PCR Technique: RT-PCR, Eaton Publishing Company/ BioTechniques Books (1995).
- 10 PCR and hybridization methods may be used to identify and/or isolate nucleic acid molecules of the present invention including allelic variants, homologous nucleic acid molecules and fragments. PCR and hybridization methods may also be used to identify, amplify and/or isolate nucleic acid molecules of the present invention that encode homologous proteins, analogs, fusion protein or muteins of the invention. Nucleic acid
15 primers as described herein can be used to prime amplification of nucleic acid molecules of the invention, using transcript-derived or genomic DNA as template.

These nucleic acid primers can also be used, for example, to prime single base extension (SBE) for SNP detection (*See, e.g.*, U.S. Pat. No. 6,004,744, the disclosure of which is incorporated herein by reference in its entirety).

- 20 Isothermal amplification approaches, such as rolling circle amplification, are also now well-described. *See, e.g.*, Schweitzer *et al.*, *Curr. Opin. Biotechnol.* 12(1): 21-7 (2001); international patent publications WO 97/19193 and WO 00/15779, and United States Patent Nos. 5,854,033 and 5,714,320, the disclosures of which are incorporated herein by reference in their entireties. Rolling circle amplification can be combined with
25 other techniques to facilitate SNP detection. *See, e.g.*, Lizardi *et al.*, *Nature Genet.* 19(3): 225-32 (1998).

- Nucleic acid molecules of the present invention may be bound to a substrate either covalently or noncovalently. The substrate can be porous or solid, planar or non-planar, unitary or distributed. The bound nucleic acid molecules may be used as hybridization
30 probes, and may be labeled or unlabeled. In a preferred embodiment, the bound nucleic acid molecules are unlabeled.

In one embodiment, the nucleic acid molecule of the present invention is bound to a porous substrate, *e.g.*, a membrane, typically comprising nitrocellulose, nylon, or

positively-charged derivatized nylon. The nucleic acid molecule of the present invention can be used to detect a hybridizing nucleic acid molecule that is present within a labeled nucleic acid sample, *e.g.*, a sample of transcript-derived nucleic acids. In another embodiment, the nucleic acid molecule is bound to a solid substrate, including, without
5 limitation, glass, amorphous silicon, crystalline silicon or plastics. Examples of plastics include, without limitation, polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof. The solid substrate may be any shape, including rectangular, disk-like
10 and spherical. In a preferred embodiment, the solid substrate is a microscope slide or slide-shaped substrate.

The nucleic acid molecule of the present invention can be attached covalently to a surface of the support substrate or applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence by presumed noncovalent interactions, or some
15 combination thereof. The nucleic acid molecule of the present invention can be bound to a substrate to which a plurality of other nucleic acids are concurrently bound, hybridization to each of the plurality of bound nucleic acids being separately detectable. At low density, *e.g.* on a porous membrane, these substrate-bound collections are typically denominated macroarrays; at higher density, typically on a solid support, such as glass, these substrate
20 bound collections of plural nucleic acids are colloquially termed microarrays. As used herein, the term microarray includes arrays of all densities. It is, therefore, another aspect of the invention to provide microarrays that comprise one or more of the nucleic acid molecules of the present invention.

In yet another embodiment, the invention is directed to single exon probes based
25 on the CSNAs disclosed herein.

Expression Vectors, Host Cells and Recombinant Methods of Producing Polypeptides

Another aspect of the present invention provides vectors that comprise one or more
30 of the isolated nucleic acid molecules of the present invention, and host cells in which such vectors have been introduced.

The vectors can be used, *inter alia*, for propagating the nucleic acid molecules of the present invention in host cells (cloning vectors), for shuttling the nucleic acid

molecules of the present invention between host cells derived from disparate organisms (shuttle vectors), for inserting the nucleic acid molecules of the present invention into host cell chromosomes (insertion vectors), for expressing sense or antisense RNA transcripts of the nucleic acid molecules of the present invention *in vitro* or within a host cell, and for
5 expressing polypeptides encoded by the nucleic acid molecules of the present invention, alone or as fusion proteins with heterologous polypeptides (expression vectors). Vectors are by now well-known in the art, and are described, *inter alia*, in Jones *et al.* (eds.), Vectors: Cloning Applications: Essential Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Jones *et al.* (eds.), Vectors: Expression Systems: Essential
10 Techniques (Essential Techniques Series), John Wiley & Son Ltd. (1998); Gacesa *et al.*, Vectors: Essential Data, John Wiley & Sons Ltd. (1995); Cid-Arregui (eds.), Viral Vectors: Basic Science and Gene Therapy, Eaton Publishing Co. (2000); Sambrook (2001), *supra*; Ausubel (1999), *supra*. Furthermore, an enormous variety of vectors are available commercially. Use of existing vectors and modifications thereof are well within
15 the skill in the art. Thus, only basic features need be described here.

Nucleic acid sequences may be expressed by operatively linking them to an expression control sequence in an appropriate expression vector and employing that expression vector to transform an appropriate unicellular host. Expression control sequences are sequences which control the transcription, post-transcriptional events and
20 translation of nucleic acid sequences. Such operative linking of a nucleic sequence of this invention to an expression control sequence, of course, includes, if not already part of the nucleic acid sequence, the provision of a translation initiation codon, ATG or GTG, in the correct reading frame upstream of the nucleic acid sequence.

A wide variety of host/expression vector combinations may be employed in
25 expressing the nucleic acid sequences of this invention. Useful expression vectors, for example, may consist of segments of chromosomal, non-chromosomal and synthetic nucleic acid sequences.

In one embodiment, prokaryotic cells may be used with an appropriate vector. Prokaryotic host cells are often used for cloning and expression. In a preferred
30 embodiment, prokaryotic host cells include *E. coli*, *Pseudomonas*, *Bacillus* and *Streptomyces*. In a preferred embodiment, bacterial host cells are used to express the nucleic acid molecules of the instant invention. Useful expression vectors for bacterial hosts include bacterial plasmids, such as those from *E. coli*, *Bacillus* or *Streptomyces*,

including pBluescript, pGEX-2T, pUC vectors, col E1, pCR1, pBR322, pMB9 and their derivatives, wider host range plasmids, such as RP4, phage DNAs, *e.g.*, the numerous derivatives of phage lambda, *e.g.*, NM989, λ GT10 and λ GT11, and other phages, *e.g.*, M13 and filamentous single stranded phage DNA. Where *E. coli* is used as host, 5 selectable markers are, analogously, chosen for selectivity in gram negative bacteria: *e.g.*, typical markers confer resistance to antibiotics, such as ampicillin, tetracycline, chloramphenicol, kanamycin, streptomycin and zeocin; auxotrophic markers can also be used.

In other embodiments, eukaryotic host cells, such as yeast, insect, mammalian or 10 plant cells, may be used. Yeast cells, typically *S. cerevisiae*, are useful for eukaryotic genetic studies, due to the ease of targeting genetic changes by homologous recombination and the ability to easily complement genetic defects using recombinantly expressed proteins. Yeast cells are useful for identifying interacting protein components, *e.g.* through use of a two-hybrid system. In a preferred embodiment, yeast cells are useful for 15 protein expression. Vectors of the present invention for use in yeast will typically, but not invariably, contain an origin of replication suitable for use in yeast and a selectable marker that is functional in yeast. Yeast vectors include Yeast Integrating plasmids (*e.g.*, YIp5) and Yeast Replicating plasmids (the YRp and YEplac series plasmids), Yeast Centromere plasmids (the YCp series plasmids), Yeast Artificial Chromosomes (YACs) which are 20 based on yeast linear plasmids, denoted YLp, pGPD-2, 2 μ plasmids and derivatives thereof, and improved shuttle vectors such as those described in Gietz *et al.*, *Gene*, 74: 527-34 (1988) (YIplac, YEplac and YCplac). Selectable markers in yeast vectors include a variety of auxotrophic markers, the most common of which are (in *Saccharomyces cerevisiae*) URA3, HIS3, LEU2, TRP1 and LYS2, which complement specific 25 auxotrophic mutations, such as *ura3-52*, *his3-D1*, *leu2-D1*, *trp1-D1* and *lys2-201*.

Insect cells are often chosen for high efficiency protein expression. Where the host cells are from *Spodoptera frugiperda*, *e.g.*, Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA), the vector replicative strategy is typically based upon the baculovirus life cycle. Typically, baculovirus transfer vectors are used to 30 replace the wild-type AcMNPV polyhedrin gene with a heterologous gene of interest. Sequences that flank the polyhedrin gene in the wild-type genome are positioned 5' and 3' of the expression cassette on the transfer vectors. Following co-transfection with AcMNPV DNA, a homologous recombination event occurs between these sequences

resulting in a recombinant virus carrying the gene of interest and the polyhedrin or p10 promoter. Selection can be based upon visual screening for lacZ fusion activity.

The host cells may also be mammalian cells, which are particularly useful for expression of proteins intended as pharmaceutical agents, and for screening of potential agonists and antagonists of a protein or a physiological pathway. Mammalian vectors intended for autonomous extrachromosomal replication will typically include a viral origin, such as the SV40 origin (for replication in cell lines expressing the large T-antigen, such as COS1 and COS7 cells), the papillomavirus origin, or the EBV origin for long term episomal replication (for use, *e.g.*, in 293-EBNA cells, which constitutively express the EBV EBNA-1 gene product and adenovirus E1A). Vectors intended for integration, and thus replication as part of the mammalian chromosome, can, but need not, include an origin of replication functional in mammalian cells, such as the SV40 origin. Vectors based upon viruses, such as adenovirus, adeno-associated virus, vaccinia virus, and various mammalian retroviruses, will typically replicate according to the viral replicative strategy. Selectable markers for use in mammalian cells include, include but are not limited to, resistance to neomycin (G418), blasticidin, hygromycin and zeocin, and selection based upon the purine salvage pathway using HAT medium.

Expression in mammalian cells can be achieved using a variety of plasmids, including pSV2, pBC12BI, and p91023, as well as lytic virus vectors (*e.g.*, vaccinia virus, adeno virus, and baculovirus), episomal virus vectors (*e.g.*, bovine papillomavirus), and retroviral vectors (*e.g.*, murine retroviruses). Useful vectors for insect cells include baculoviral vectors and pVL 941.

Plant cells can also be used for expression, with the vector replicon typically derived from a plant virus (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) and selectable markers chosen for suitability in plants.

It is known that codon usage of different host cells may be different. For example, a plant cell and a human cell may exhibit a difference in codon preference for encoding a particular amino acid. As a result, human mRNA may not be efficiently translated in a plant, bacteria or insect host cell. Therefore, another embodiment of this invention is directed to codon optimization. The codons of the nucleic acid molecules of the invention may be modified to resemble, as much as possible, genes naturally contained within the host cell without altering the amino acid sequence encoded by the nucleic acid molecule.

Any of a wide variety of expression control sequences may be used in these vectors to express the nucleic acid molecules of this invention. Such useful expression control sequences include the expression control sequences associated with structural genes of the foregoing expression vectors. Expression control sequences that control transcription include, *e.g.*, promoters, enhancers and transcription termination sites. Expression control sequences in eukaryotic cells that control post-transcriptional events include splice donor and acceptor sites and sequences that modify the half-life of the transcribed RNA, *e.g.*, sequences that direct poly(A) addition or binding sites for RNA-binding proteins. Expression control sequences that control translation include ribosome binding sites, sequences which direct targeted expression of the polypeptide to or within particular cellular compartments, and sequences in the 5' and 3' untranslated regions that modify the rate or efficiency of translation.

Examples of useful expression control sequences for a prokaryote, *e.g.*, *E. coli*, will include a promoter, often a phage promoter, such as phage lambda pL promoter, the trc promoter, a hybrid derived from the trp and lac promoters, the bacteriophage T7 promoter (in *E. coli* cells engineered to express the T7 polymerase), the TAC or TRC system, the major operator and promoter regions of phage lambda, the control regions of fd coat protein, and the araBAD operon. Prokaryotic expression vectors may further include transcription terminators, such as the aspA terminator, and elements that facilitate translation, such as a consensus ribosome binding site and translation termination codon, Schomer *et al.*, *Proc. Natl. Acad. Sci. USA* 83: 8506-8510 (1986).

Expression control sequences for yeast cells, typically *S. cerevisiae*, will include a yeast promoter, such as the CYC1 promoter, the GAL1 promoter, the GAL10 promoter, ADH1 promoter, the promoters of the yeast α -mating system, or the GPD promoter, and will typically have elements that facilitate transcription termination, such as the transcription termination signals from the CYC1 or ADH1 gene.

Expression vectors useful for expressing proteins in mammalian cells will include a promoter active in mammalian cells. These promoters include, but are not limited to, those derived from mammalian viruses, such as the enhancer-promoter sequences from the immediate early gene of the human cytomegalovirus (CMV), the enhancer-promoter sequences from the Rous sarcoma virus long terminal repeat (RSV LTR), the enhancer-promoter from SV40 and the early and late promoters of adenovirus. Other expression control sequences include the promoter for 3-phosphoglycerate kinase or other glycolytic

enzymes, the promoters of acid phosphatase. Other expression control sequences include those from the gene comprising the CSNA of interest. Often, expression is enhanced by incorporation of polyadenylation sites, such as the late SV40 polyadenylation site and the polyadenylation signal and transcription termination sequences from the bovine growth hormone (BGH) gene, and ribosome binding sites. Furthermore, vectors can include introns, such as intron II of rabbit β -globin gene and the SV40 splice elements.

Preferred nucleic acid vectors also include a selectable or amplifiable marker gene and means for amplifying the copy number of the gene of interest. Such marker genes are well-known in the art. Nucleic acid vectors may also comprise stabilizing sequences (*e.g.*, ori- or ARS-like sequences and telomere-like sequences), or may alternatively be designed to favor directed or non-directed integration into the host cell genome. In a preferred embodiment, nucleic acid sequences of this invention are inserted in frame into an expression vector that allows high level expression of an RNA which encodes a protein comprising the encoded nucleic acid sequence of interest. Nucleic acid cloning and sequencing methods are well-known to those of skill in the art and are described in an assortment of laboratory manuals, including Sambrook (1989), *supra*, Sambrook (2000), *supra*; and Ausubel (1992), *supra*, Ausubel (1999), *supra*. Product information from manufacturers of biological, chemical and immunological reagents also provide useful information.

Expression vectors may be either constitutive or inducible. Inducible vectors include either naturally inducible promoters, such as the *trc* promoter, which is regulated by the *lac* operon, and the *pL* promoter, which is regulated by tryptophan, the MMTV-LTR promoter, which is inducible by dexamethasone, or can contain synthetic promoters and/or additional elements that confer inducible control on adjacent promoters. Examples of inducible synthetic promoters are the hybrid *Plac/ara-1* promoter and the *PLtetO-1* promoter. The *PLtetO-1* promoter takes advantage of the high expression levels from the *PL* promoter of phage lambda, but replaces the lambda repressor sites with two copies of operator 2 of the *Tn10* tetracycline resistance operon, causing this promoter to be tightly repressed by the Tet repressor protein and induced in response to tetracycline (Tc) and Tc derivatives such as anhydrotetracycline. Vectors may also be inducible because they contain hormone response elements, such as the glucocorticoid response element (GRE) and the estrogen response element (ERE), which can confer hormone inducibility where vectors are used for expression in cells having the respective hormone

receptors. To reduce background levels of expression, elements responsive to ecdysone, an insect hormone, can be used instead, with coexpression of the ecdysone receptor.

In one embodiment of the invention, expression vectors can be designed to fuse the expressed polypeptide to small protein tags that facilitate purification and/or visualization.

- 5 Such tags include a polyhistidine tag that facilitates purification of the fusion protein by immobilized metal affinity chromatography, for example using NiNTA resin (Qiagen Inc., Valencia, CA, USA) or TALON™ resin (cobalt immobilized affinity chromatography medium, Clontech Labs, Palo Alto, CA, USA). The fusion protein can include a chitin-binding tag and self-excising intein, permitting chitin-based purification with self-removal
- 10 of the fused tag (IMPACT™ system, New England Biolabs, Inc., Beverly, MA, USA). Alternatively, the fusion protein can include a calmodulin-binding peptide tag, permitting purification by calmodulin affinity resin (Stratagene, La Jolla, CA, USA), or a specifically excisable fragment of the biotin carboxylase carrier protein, permitting purification of *in vivo* biotinylated protein using an avidin resin and subsequent tag removal (Promega,
- 15 Madison, WI, USA). As another useful alternative, the polypeptides of the present invention can be expressed as a fusion to glutathione-S-transferase, the affinity and specificity of binding to glutathione permitting purification using glutathione affinity resins, such as Glutathione-Superflow Resin (Clontech Laboratories, Palo Alto, CA, USA), with subsequent elution with free glutathione. Other tags include, for example, the
- 20 Xpress epitope, detectable by anti-Xpress antibody (Invitrogen, Carlsbad, CA, USA), a myc tag, detectable by anti-myc tag antibody, the V5 epitope, detectable by anti-V5 antibody (Invitrogen, Carlsbad, CA, USA), FLAG® epitope, detectable by anti-FLAG® antibody (Stratagene, La Jolla, CA, USA), and the HA epitope, detectable by anti-HA antibody.

- 25 For secretion of expressed polypeptides, vectors can include appropriate sequences that encode secretion signals, such as leader peptides. For example, the pSecTag2 vectors (Invitrogen, Carlsbad, CA, USA) are 5.2 kb mammalian expression vectors that carry the secretion signal from the V-J2-C region of the mouse Ig kappa-chain for efficient secretion of recombinant proteins from a variety of mammalian cell lines.

- 30 Expression vectors can also be designed to fuse proteins encoded by the heterologous nucleic acid insert to polypeptides that are larger than purification and/or identification tags. Useful protein fusions include those that permit display of the encoded protein on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such

as those that have a green fluorescent protein (GFP)-like chromophore, fusions to the IgG Fc region, and fusions for use in two hybrid systems.

- Vectors for phage display fuse the encoded polypeptide to, *e.g.*, the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13. *See* Barbas *et al.*, Phage Display: A Laboratory Manual, Cold Spring Harbor Laboratory Press (2001); Kay *et al.* (eds.), Phage Display of Peptides and Proteins: A Laboratory Manual, Academic Press, Inc., (1996); Abelson *et al.* (eds.), Combinatorial Chemistry (Methods in Enzymology, Vol. 267) Academic Press (1996). Vectors for yeast display, *e.g.* the pYD1 yeast display vector (Invitrogen, Carlsbad, CA, USA), use the α -agglutinin yeast adhesion receptor to display recombinant protein on the surface of *S. cerevisiae*. Vectors for mammalian display, *e.g.*, the pDisplay™ vector (Invitrogen, Carlsbad, CA, USA), target recombinant proteins using an N-terminal cell surface targeting signal and a C-terminal transmembrane anchoring domain of platelet derived growth factor receptor.
- A wide variety of vectors now exist that fuse proteins encoded by heterologous nucleic acids to the chromophore of the substrate-independent, intrinsically fluorescent green fluorescent protein from *Aequorea victoria* ("GFP") and its variants. The GFP-like chromophore can be selected from GFP-like chromophores found in naturally occurring proteins, such as *A. victoria* GFP (GenBank accession number AAA27721), *Renilla reniformis* GFP, FP583 (GenBank accession no. AF168419) (DsRed), FP593 (AF272711), FP483 (AF168420), FP484 (AF168424), FP595 (AF246709), FP486 (AF168421), FP538 (AF168423), and FP506 (AF168422), and need include only so much of the native protein as is needed to retain the chromophore's intrinsic fluorescence. Methods for determining the minimal domain required for fluorescence are known in the art. *See* Li *et al.*, *J. Biol. Chem.* 272: 28545-28549 (1997). Alternatively, the GFP-like chromophore can be selected from GFP-like chromophores modified from those found in nature. The methods for engineering such modified GFP-like chromophores and testing them for fluorescence activity, both alone and as part of protein fusions, are well-known in the art. *See* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Palm *et al.*, *Methods Enzymol.* 302: 378-394 (1999).
- A variety of such modified chromophores are now commercially available and can readily be used in the fusion proteins of the present invention. These include EGFP ("enhanced GFP"), EBFP ("enhanced blue fluorescent protein"), BFP2, EYFP ("enhanced yellow fluorescent protein"), ECFP ("enhanced cyan fluorescent protein") or Citrine. EGFP (*see*,

e.g. Cormack *et al.*, *Gene* 173: 33–38 (1996); United States Patent Nos. 6,090,919 and 5,804,387) is found on a variety of vectors, both plasmid and viral, which are available commercially (Clontech Labs, Palo Alto, CA, USA); EBFP is optimized for expression in mammalian cells whereas BFP2, which retains the original jellyfish codons, can be expressed in bacteria (*see, e.g.* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Cormack *et al.*, *Gene* 173: 33-38 (1996)). Vectors containing these blue-shifted variants are available from Clontech Labs (Palo Alto, CA, USA). Vectors containing EYFP, ECFP (*see, e.g.* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996); Miyawaki *et al.*, *Nature* 388: 882-887 (1997)) and Citrine (*see, e.g.* Heikal *et al.*, *Proc. Natl. Acad. Sci. USA* 97: 11996-12001 (2000)) are also available from Clontech Labs. The GFP-like chromophore can also be drawn from other modified GFPs, including those described in United States Patent Nos. 6,124,128; 6,096,865; 6,090,919; 6,066,476; 6,054,321; 6,027,881; 5,968,750; 5,874,304; 5,804,387; 5,777,079; 5,741,668; and 5,625,048, the disclosures of which are incorporated herein by reference in their entireties. *See also* Conn (ed.), Green Fluorescent Protein (Methods in Enzymology, Vol. 302), Academic Press, Inc. (1999). The GFP-like chromophore of each of these GFP variants can usefully be included in the fusion proteins of the present invention.

Fusions to the IgG Fc region increase serum half life of protein pharmaceutical products through interaction with the FcRn receptor (also denominated the FcRp receptor and the Brambell receptor, FcRb), further described in International Patent Application nos. WO 97/43316, WO 97/34631, WO 96/32478, WO 96/18412.

For long-term, high-yield recombinant production of the polypeptides of the present invention, stable expression is preferred. Stable expression is readily achieved by integration into the host cell genome of vectors having selectable markers, followed by selection of these integrants. Vectors such as pUB6/V5-His A, B, and C (Invitrogen, Carlsbad, CA, USA) are designed for high-level stable expression of heterologous proteins in a wide range of mammalian tissue types and cell lines. pUB6/V5-His uses the promoter/enhancer sequence from the human ubiquitin C gene to drive expression of recombinant proteins: expression levels in 293, CHO, and NIH3T3 cells are comparable to levels from the CMV and human EF-1a promoters. The *bsd* gene permits rapid selection of stably transfected mammalian cells with the potent antibiotic blasticidin.

Replication incompetent retroviral vectors, typically derived from Moloney murine leukemia virus, also are useful for creating stable transfectants having integrated provirus.

The highly efficient transduction machinery of retroviruses, coupled with the availability of a variety of packaging cell lines such as RetroPack™ PT 67, EcoPack2™-293, AmphoPack-293, and GP2-293 cell lines (all available from Clontech Laboratories, Palo Alto, CA, USA) allow a wide host range to be infected with high efficiency; varying the
5 multiplicity of infection readily adjusts the copy number of the integrated provirus.

Of course, not all vectors and expression control sequences will function equally well to express the nucleic acid molecules of this invention. Neither will all hosts function equally well with the same expression system. However, one of skill in the art may make a selection among these vectors, expression control sequences and hosts without undue
10 experimentation and without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must be replicated in it. The vector's copy number, the ability to control that copy number, the ability to control integration, if any, and the expression of any other proteins encoded by the vector, such as antibiotic or other selection markers, should also be considered. The present invention
15 further includes host cells comprising the vectors of the present invention, either present episomally within the cell or integrated, in whole or in part, into the host cell chromosome. Among other considerations, some of which are described above, a host cell strain may be chosen for its ability to process the expressed polypeptide in the desired fashion. Such post-translational modifications of the polypeptide include, but are not limited to,
20 acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation, and it is an aspect of the present invention to provide CSPs with such post-translational modifications.

In selecting an expression control sequence, a variety of factors should also be considered. These include, for example, the relative strength of the sequence, its
25 controllability, and its compatibility with the nucleic acid molecules of this invention, particularly with regard to potential secondary structures. Unicellular hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity of the product coded for by the nucleic acid sequences of this invention, their secretion characteristics, their ability to fold the polypeptide correctly, their fermentation or culture
30 requirements, and the ease of purification from them of the products coded for by the nucleic acid molecules of this invention.

The recombinant nucleic acid molecules and more particularly, the expression vectors of this invention may be used to express the polypeptides of this invention as

recombinant polypeptides in a heterologous host cell. The polypeptides of this invention may be full-length or less than full-length polypeptide fragments recombinantly expressed from the nucleic acid molecules according to this invention. Such polypeptides include analogs, derivatives and muteins that may or may not have biological activity.

5 Vectors of the present invention will also often include elements that permit *in vitro* transcription of RNA from the inserted heterologous nucleic acid. Such vectors typically include a phage promoter, such as that from T7, T3, or SP6, flanking the nucleic acid insert. Often two different such promoters flank the inserted nucleic acid, permitting separate *in vitro* production of both sense and antisense strands.

10 Transformation and other methods of introducing nucleic acids into a host cell (e.g., conjugation, protoplast transformation or fusion, transfection, electroporation, liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion) can be accomplished by a variety of methods which are well-known in the art (See, for instance, Ausubel, *supra*, and Sambrook *et al.*, *supra*).

15 Bacterial, yeast, plant or mammalian cells are transformed or transfected with an expression vector, such as a plasmid, a cosmid, or the like, wherein the expression vector comprises the nucleic acid of interest. Alternatively, the cells may be infected by a viral expression vector comprising the nucleic acid of interest. Depending upon the host cell, vector, and method of transformation used, transient or stable expression of the
20 polypeptide will be constitutive or inducible. One having ordinary skill in the art will be able to decide whether to express a polypeptide transiently or stably, and whether to express the protein constitutively or inducibly.

A wide variety of unicellular host cells are useful in expressing the DNA sequences of this invention. These hosts may include well-known eukaryotic and
25 prokaryotic hosts, such as strains of, fungi, yeast, insect cells such as *Spodoptera frugiperda* (SF9), animal cells such as CHO, as well as plant cells in tissue culture. Representative examples of appropriate host cells include, but are not limited to, bacterial cells, such as *E. coli*, *Caulobacter crescentus*, *Streptomyces* species, and *Salmonella typhimurium*; yeast cells, such as *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*,
30 *Pichia pastoris*, *Pichia methanolica*; insect cell lines, such as those from *Spodoptera frugiperda* — e.g., Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA) — *Drosophila* S2 cells, and *Trichoplusia ni* High Five® Cells (Invitrogen, Carlsbad, CA, USA); and mammalian cells. Typical mammalian cells include

BHK cells, BSC 1 cells, BSC 40 cells, BMT 10 cells, VERO cells, COS1 cells, COS7 cells, Chinese hamster ovary (CHO) cells, 3T3 cells, NIH 3T3 cells, 293 cells, HEPG2 cells, HeLa cells, L cells, MDCK cells, HEK293 cells, WI38 cells, murine ES cell lines (e.g., from strains 129/SV, C57/BL6, DBA-1, 129/SVJ), K562 cells, Jurkat cells, and
5 BW5147 cells. Other mammalian cell lines are well-known and readily available from the American Type Culture Collection (ATCC) (Manassas, VA, USA) and the National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Cell Repositories (Camden, NJ, USA). Cells or cell lines derived from colon are particularly preferred because they may provide a more native post-translational
10 processing. Particularly preferred are human colon cells.

Particular details of the transfection, expression and purification of recombinant proteins are well documented and are understood by those of skill in the art. Further details on the various technical aspects of each of the steps used in recombinant production of foreign genes in bacterial cell expression systems can be found in a number
15 of texts and laboratory manuals in the art. See, e.g., Ausubel (1992), *supra*, Ausubel (1999), *supra*, Sambrook (1989), *supra*, and Sambrook (2001), *supra*.

Methods for introducing the vectors and nucleic acid molecules of the present invention into the host cells are well-known in the art; the choice of technique will depend primarily upon the specific vector to be introduced and the host cell chosen.

20 Nucleic acid molecules and vectors may be introduced into prokaryotes, such as *E. coli*, in a number of ways. For instance, phage lambda vectors will typically be packaged using a packaging extract (e.g., Gigapack® packaging extract, Stratagene, La Jolla, CA, USA), and the packaged virus used to infect *E. coli*.

Plasmid vectors will typically be introduced into chemically competent or
25 electrocompetent bacterial cells. *E. coli* cells can be rendered chemically competent by treatment, e.g., with CaCl₂, or a solution of Mg²⁺, Mn²⁺, Ca²⁺, Rb⁺ or K⁺, dimethyl sulfoxide, dithiothreitol, and hexamine cobalt (III), Hanahan, *J. Mol. Biol.* 166(4):557-80 (1983), and vectors introduced by heat shock. A wide variety of chemically competent strains are also available commercially (e.g., Epicurian Coli® XL10-Gold®
30 Ultracompetent Cells (Stratagene, La Jolla, CA, USA); DH5α competent cells (Clontech Laboratories, Palo Alto, CA, USA); and TOP10 Chemically Competent *E. coli* Kit (Invitrogen, Carlsbad, CA, USA)). Bacterial cells can be rendered electrocompetent to take up exogenous DNA by electroporation by various pre-pulse treatments; vectors are

introduced by electroporation followed by subsequent outgrowth in selected media. An extensive series of protocols is provided by BioRad (Richmond, CA, USA).

Vectors can be introduced into yeast cells by spheroplasting, treatment with lithium salts, electroporation, or protoplast fusion. Spheroplasts are prepared by the action of hydrolytic enzymes such as a snail-gut extract, usually denoted Glusulase or Zymolyase, or an enzyme from *Arthrobacter luteus* to remove portions of the cell wall in the presence of osmotic stabilizers, typically 1 M sorbitol. DNA is added to the spheroplasts, and the mixture is co-precipitated with a solution of polyethylene glycol (PEG) and Ca^{2+} . Subsequently, the cells are resuspended in a solution of sorbitol, mixed with molten agar and then layered on the surface of a selective plate containing sorbitol.

For lithium-mediated transformation, yeast cells are treated with lithium acetate to permeabilize the cell wall, DNA is added and the cells are co-precipitated with PEG. The cells are exposed to a brief heat shock, washed free of PEG and lithium acetate, and subsequently spread on plates containing ordinary selective medium. Increased frequencies of transformation are obtained by using specially-prepared single-stranded carrier DNA and certain organic solvents. Schiestl *et al.*, *Curr. Genet.* 16(5-6): 339-46 (1989).

For electroporation, freshly-grown yeast cultures are typically washed, suspended in an osmotic protectant, such as sorbitol, mixed with DNA, and the cell suspension pulsed in an electroporation device. Subsequently, the cells are spread on the surface of plates containing selective media. Becker *et al.*, *Methods Enzymol.* 194: 182-187 (1991). The efficiency of transformation by electroporation can be increased over 100-fold by using PEG, single-stranded carrier DNA and cells that are in late log-phase of growth. Larger constructs, such as YACs, can be introduced by protoplast fusion.

Mammalian and insect cells can be directly infected by packaged viral vectors, or transfected by chemical or electrical means. For chemical transfection, DNA can be coprecipitated with CaPO_4 or introduced using liposomal and nonliposomal lipid-based agents. Commercial kits are available for CaPO_4 transfection (CalPhos™ Mammalian Transfection Kit, Clontech Laboratories, Palo Alto, CA, USA), and lipid-mediated transfection can be practiced using commercial reagents, such as LIPOFECTAMINE™ 2000, LIPOFECTAMINE™ Reagent, CELLFECTIN® Reagent, and LIPOFECTIN® Reagent (Invitrogen, Carlsbad, CA, USA), DOTAP Liposomal Transfection Reagent, FuGENE 6, X-tremeGENE Q2, DOSPER, (Roche Molecular Biochemicals, Indianapolis,

IN USA), Effectene™, PolyFect®, Superfect® (Qiagen, Inc., Valencia, CA, USA).

Protocols for electroporating mammalian cells can be found in, for example, ; Norton *et al.* (eds.), Gene Transfer Methods: Introducing DNA into Living Cells and Organisms, BioTechniques Books, Eaton Publishing Co. (2000). Other transfection techniques

- 5 include transfection by particle bombardment and microinjection. *See, e.g., Cheng et al., Proc. Natl. Acad. Sci. USA* 90(10): 4455-9 (1993); Yang *et al., Proc. Natl. Acad. Sci. USA* 87(24): 9568-72 (1990).

Production of the recombinantly produced proteins of the present invention can optionally be followed by purification.

- 10 Purification of recombinantly expressed proteins is now well within the skill in the art and thus need not be detailed here. *See, e.g., Thorner et al.* (eds.), Applications of Chimeric Genes and Hybrid Proteins, Part A: Gene Expression and Protein Purification (Methods in Enzymology, Vol. 326), Academic Press (2000); Harbin (ed.), Cloning, Gene Expression and Protein Purification : Experimental Procedures and Process Rationale,
15 Oxford Univ. Press (2001); Marshak *et al., Strategies for Protein Purification and Characterization: A Laboratory Course Manual*, Cold Spring Harbor Laboratory Press (1996); and Roe (ed.), Protein Purification Applications, Oxford University Press (2001).

- Briefly, however, if purification tags have been fused through use of an expression vector that appends such tag, purification can be effected, at least in part, by means
20 appropriate to the tag, such as use of immobilized metal affinity chromatography for polyhistidine tags. Other techniques common in the art include ammonium sulfate fractionation, immunoprecipitation, fast protein liquid chromatography (FPLC), high performance liquid chromatography (HPLC), and preparative gel electrophoresis.

- Polypeptides, including Fragments Muteins, Homologous Proteins, Allelic Variants,
25 Analogs and Derivatives

- Another aspect of the invention relates to polypeptides encoded by the nucleic acid molecules described herein. In a preferred embodiment, the polypeptide is a colon specific polypeptide (CSP). In an even more preferred embodiment, the polypeptide comprises an amino acid sequence of SEQ ID NO:59-82 or is derived from a polypeptide
30 having the amino acid sequence of SEQ ID NO: 101-190. A polypeptide as defined herein may be produced recombinantly, as discussed *supra*, may be isolated from a cell that naturally expresses the protein, or may be chemically synthesized following the teachings

of the specification and using methods well-known to those having ordinary skill in the art.

Polypeptides of the present invention may also comprise a part or fragment of a CSP. In a preferred embodiment, the fragment is derived from a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO: 101-190.

Polypeptides of the present invention comprising a part or fragment of an entire CSP may or may not be CSPs. For example, a full-length polypeptide may be colon-specific, while a fragment thereof may be found in other tissues as well as in colon. A polypeptide that is not a CSP, whether it is a fragment, analog, mutein, homologous protein or derivative, is nevertheless useful, especially for immunizing animals to prepare anti-CSP antibodies. In a preferred embodiment, the part or fragment is a CSP. Methods of determining whether a polypeptide of the present invention is a CSP are described *infra*.

Polypeptides of the present invention comprising fragments of at least 6 contiguous amino acids are also useful in mapping B cell and T cell epitopes of the reference protein. *See, e.g., Geysen et al., Proc. Natl. Acad. Sci. USA* 81: 3998-4002 (1984) and United States Patent Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. Because the fragment need not itself be immunogenic, part of an immunodominant epitope, nor even recognized by native antibody, to be useful in such epitope mapping, all fragments of at least 6 amino acids of a polypeptide of the present invention have utility in such a study.

Polypeptides of the present invention comprising fragments of at least 8 contiguous amino acids, often at least 15 contiguous amino acids, are useful as immunogens for raising antibodies that recognize polypeptides of the present invention. *See, e.g., Lerner, Nature* 299: 592-596 (1982); Shinnick *et al., Annu. Rev. Microbiol.* 37: 425-46 (1983); Sutcliffe *et al., Science* 219: 660-6 (1983). As further described in the above-cited references, virtually all 8-mers, conjugated to a carrier, such as a protein, prove immunogenic and are capable of eliciting antibody for the conjugated peptide; accordingly, all fragments of at least 8 amino acids of the polypeptides of the present invention have utility as immunogens.

Polypeptides comprising fragments of at least 8, 9, 10 or 12 contiguous amino acids are also useful as competitive inhibitors of binding of the entire polypeptide, or a portion thereof, to antibodies (as in epitope mapping), and to natural binding partners, such as subunits in a multimeric complex or to receptors or ligands of the subject protein;

this competitive inhibition permits identification and separation of molecules that bind specifically to the polypeptide of interest. See United States Patent Nos. 5,539,084 and 5,783,674, incorporated herein by reference in their entireties.

The polypeptide of the present invention thus preferably is at least 6 amino acids
5 in length, typically at least 8, 9, 10 or 12 amino acids in length, and often at least 15 amino acids in length. Often, the polypeptide of the present invention is at least 20 amino acids in length, even 25 amino acids, 30 amino acids, 35 amino acids, or 50 amino acids or more in length. Of course, larger polypeptides having at least 75 amino acids, 100 amino acids, or even 150 amino acids are also useful, and at times preferred.

10 One having ordinary skill in the art can produce fragments by truncating the nucleic acid molecule, *e.g.*, a CSNA, encoding the polypeptide and then expressing it recombinantly. Alternatively, one can produce a fragment by chemically synthesizing a portion of the full-length polypeptide. One may also produce a fragment by enzymatically cleaving either a recombinant polypeptide or an isolated naturally-occurring polypeptide.
15 Methods of producing polypeptide fragments are well-known in the art. *See, e.g.*, Sambrook (1989), *supra*; Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. In one embodiment, a polypeptide comprising only a fragment, preferably a fragment of a CSP, may be produced by chemical or enzymatic cleavage of a CSP polypeptide. In a preferred embodiment, a polypeptide fragment is produced by
20 expressing a nucleic acid molecule of the present invention encoding a fragment, preferably of a CSP, in a host cell.

Polypeptides of the present invention are also inclusive of mutants, fusion proteins, homologous proteins and allelic variants.

A mutant protein, or mutein, may have the same or different properties compared
25 to a naturally-occurring polypeptide and comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of a native polypeptide. Small deletions and insertions can often be found that do not alter the function of a protein. Muteins may or may not be colon-specific. Preferably, the mutein is colon-specific. More preferably the mutein is a polypeptide that comprises at
30 least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of SEQ ID NO: 101-190. Accordingly, in a preferred embodiment, the mutein is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more

preferably at least 80% sequence identity to a CSP comprising an amino acid sequence of SEQ ID NO: 101-190. In a yet more preferred embodiment, the mutein exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97%, 98%, 99% or 99.5% sequence identity to a CSP comprising an amino acid
5 sequence of SEQ ID NO: 101-190.

A mutein may be produced by isolation from a naturally-occurring mutant cell, tissue or organism. A mutein may be produced by isolation from a cell, tissue or organism that has been experimentally mutagenized. Alternatively, a mutein may be produced by chemical manipulation of a polypeptide, such as by altering the amino acid residue to
10 another amino acid residue using synthetic or semi-synthetic chemical techniques. In a preferred embodiment, a mutein is produced from a host cell comprising a mutated nucleic acid molecule compared to the naturally-occurring nucleic acid molecule. For instance, one may produce a mutein of a polypeptide by introducing one or more mutations into a nucleic acid molecule of the invention and then expressing it recombinantly. These
15 mutations may be targeted, in which particular encoded amino acids are altered, or may be untargeted, in which random encoded amino acids within the polypeptide are altered. Muteins with random amino acid alterations can be screened for a particular biological activity or property, particularly whether the polypeptide is colon-specific, as described below. Multiple random mutations can be introduced into the gene by methods well-
20 known to the art, *e.g.*, by error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo* mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis and site-specific mutagenesis. Methods of producing muteins with targeted or random amino acid alterations are well-known in the art. *See, e.g.*, Sambrook (1989), *supra*; Sambrook
25 (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), as well as United States Patent No. 5,223,408, which is herein incorporated by reference in its entirety.

The invention also contemplates polypeptides that are homologous to a polypeptide of the invention. In a preferred embodiment, the polypeptide is homologous to a CSP. In an even more preferred embodiment, the polypeptide is homologous to a
30 CSP selected from the group having an amino acid sequence of SEQ ID NO: 101-190. By homologous polypeptide it is meant one that exhibits significant sequence identity to a CSP, preferably a CSP having an amino acid sequence of SEQ ID NO: 101-190. By significant sequence identity it is meant that the homologous polypeptide exhibits at least

50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a CSP comprising an amino acid sequence of SEQ ID NO: 101-190. More preferred are homologous polypeptides exhibiting at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97% or 98% sequence identity to a CSP comprising an amino acid sequence of SEQ ID NO: 101-190. Most preferably, the homologous polypeptide exhibits at least 99%, more preferably 99.5%, even more preferably 99.6%, 99.7%, 99.8% or 99.9% sequence identity to a CSP comprising an amino acid sequence of SEQ ID NO: 101-190. In a preferred embodiment, the amino acid substitutions of the homologous polypeptide are conservative amino acid substitutions as discussed above.

Homologous polypeptides of the present invention also comprise polypeptide encoded by a nucleic acid molecule that selectively hybridizes to a CSNA or an antisense sequence thereof. In this embodiment, it is preferred that the homologous polypeptide be encoded by a nucleic acid molecule that hybridizes to a CSNA under low stringency, moderate stringency or high stringency conditions, as defined herein. More preferred is a homologous polypeptide encoded by a nucleic acid sequence which hybridizes to a CSNA selected from the group consisting of SEQ ID NO: 1-100 or a homologous polypeptide encoded by a nucleic acid molecule that hybridizes to a nucleic acid molecule that encodes a CSP, preferably an CSP of SEQ ID NO:59-82 under low stringency, moderate stringency or high stringency conditions, as defined herein.

Homologous polypeptides of the present invention may be naturally-occurring and derived from another species, especially one derived from another primate, such as chimpanzee, gorilla, rhesus macaque, or baboon, wherein the homologous polypeptide comprises an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 101-190. The homologous polypeptide may also be a naturally-occurring polypeptide from a human, when the CSP is a member of a family of polypeptides. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, guinea pig, hamster, cow, horse, goat or pig. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring homologous protein may be isolated directly from humans or other species. Alternatively, the nucleic acid molecule

encoding the naturally-occurring homologous polypeptide may be isolated and used to express the homologous polypeptide recombinantly. The homologous polypeptide may also be one that is experimentally produced by random mutation of a nucleic acid molecule and subsequent expression of the nucleic acid molecule. Alternatively, the
5 homologous polypeptide may be one that is experimentally produced by directed mutation of one or more codons to alter the encoded amino acid of a CSP. In a preferred embodiment, the homologous polypeptide encodes a polypeptide that is a CSP.

Relatedness of proteins can also be characterized using a second functional test, the ability of a first protein competitively to inhibit the binding of a second protein to an
10 antibody. It is, therefore, another aspect of the present invention to provide isolated polypeptide not only identical in sequence to those described with particularity herein, but also to provide isolated polypeptide ("cross-reactive proteins") that competitively inhibit the binding of antibodies to all or to a portion of various of the isolated polypeptides of the present invention. Such competitive inhibition can readily be determined using
15 immunoassays well-known in the art.

As discussed above, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes, and the sequence determined from one individual of a species may differ from other allelic forms present within the population. Thus, polypeptides of the present invention are also inclusive of those encoded by an allelic variant of a nucleic acid
20 molecule encoding a CSP. In this embodiment, it is preferred that the polypeptide be encoded by an allelic variant of a gene that encodes a polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO: 101-190. More preferred is that the polypeptide be encoded by an allelic variant of a gene that has the nucleic acid sequence selected from the group consisting of SEQ ID NO: 1-100.

25 Polypeptides of the present invention are also inclusive of derivative polypeptides encoded by a nucleic acid molecule according to the instant invention. In this embodiment, it is preferred that the polypeptide be a CSP. Also preferred are derivative polypeptides having an amino acid sequence selected from the group consisting of SEQ ID NO: 101-190 and which has been acetylated, carboxylated, phosphorylated,
30 glycosylated or ubiquitinated. In another preferred embodiment, the derivative has been labeled with, *e.g.*, radioactive isotopes such as ^{125}I , ^{32}P , ^{35}S , and ^3H . In another preferred embodiment, the derivative has been labeled with fluorophores, chemiluminescent agents,

enzymes, and antigens that can serve as specific binding pair members for a labeled ligand.

Polypeptide modifications are well-known to those of skill and have been described in great detail in the scientific literature. Several particularly common
 5 modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance Creighton, Protein Structure and Molecular Properties, 2nd ed., W. H. Freeman and Company (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, in Johnson (ed.),
 10 Posttranslational Covalent Modification of Proteins, pgs. 1-12, Academic Press (1983); Seifter *et al.*, *Meth. Enzymol.* 182: 626-646 (1990) and Rattan *et al.*, *Ann. N.Y. Acad. Sci.* 663: 48-62 (1992).

One may determine whether a polypeptide of the invention is likely to be post-translationally modified by analyzing the sequence of the polypeptide to determine if there
 15 are peptide motifs indicative of sites for post-translational modification. There are a number of computer programs that permit prediction of post-translational modifications. See, e.g., www.expasy.org (accessed August 31, 2001), which includes PSORT, for prediction of protein sorting signals and localization sites, SignalP, for prediction of signal peptide cleavage sites, MITOPROT and Predotar, for prediction of mitochondrial targeting
 20 sequences, NetOGlyc, for prediction of type O-glycosylation sites in mammalian proteins, big-PI Predictor and DGPI, for prediction of prenylation-anchor and cleavage sites, and NetPhos, for prediction of Ser, Thr and Tyr phosphorylation sites in eukaryotic proteins. Other computer programs, such as those included in GCG, also may be used to determine post-translational modification peptide motifs.

25 General examples of types of post-translational modifications include, but are not limited to: (Z)-dehydrobutyric; 1-chondroitin sulfate-L-aspartic acid ester; 1'-glycosyl-L-tryptophan; 1'-phospho-L-histidine; 1-thioglycine; 2'-(S-L-cysteinyl)-L-histidine; 2'-[3-carboxamido (trimethylammonio)propyl]-L-histidine; 2'-alpha-mannosyl-L-tryptophan; 2-methyl-L-glutamine; 2-oxobutanoic acid; 2-pyrrolidone carboxylic acid; 3'-(1'-L-histidyl)-
 30 L-tyrosine; 3'-(8alpha-FAD)-L-histidine; 3'-(S-L-cysteinyl)-L-tyrosine; 3', 3'',5'-triiodo-L-thyronine; 3'-4'-phospho-L-tyrosine; 3-hydroxy-L-proline; 3'-methyl-L-histidine; 3-methyl-L-lanthionine; 3'-phospho-L-histidine; 4'-(L-tryptophan)-L-tryptophyl quinone; 42 N-cysteinyl-glycosylphosphatidylinositol ethanolamine; 43 -(T-L-histidyl)-L-tyrosine; 4-

- hydroxy-L-arginine; 4-hydroxy-L-lysine; 4-hydroxy-L-proline; 5'-(N6-L-lysine)-L-topaquinone; 5-hydroxy-L-lysine; 5-methyl-L-arginine; alpha-l-microglobulin-Ig alpha complex chromophore; bis-L-cysteinyl bis-L-histidino diron disulfide; bis-L--cysteinyl-L-N3'-histidino-L-serinyl tetrairon' tetrasulfide; chondroitin sulfate D-glucuronyl-D-galactosyl-D-galactosyl-D-xylosyl-L-serine; D-alanine; D-allo-isoleucine; D-asparagine; 5 dehydroalanine; dehydrotyrosine; dermatan 4-sulfate D-glucuronyl-D-galactosyl-D-galactosyl-D-xylosyl-L-serine; D-glucuronyl-N-glycine; dipyrrolylmethanemethyl-L-cysteine; D-leucine; D-methionine; D-phenylalanine; D-serine; D-tryptophan; glycine amide; glycine oxazolecarboxylic acid; glycine thiazolecarboxylic acid; heme P450-bis-L-10 cysteine-L-tyrosine; heme-bis-L-cysteine; hemediol-L-aspartyl ester-L-glutamyl ester; hemediol-L-aspartyl ester-L-glutamyl ester-L-methionine sulfonium; heme-L-cysteine; heme-L-histidine; heparan sulfate D-glucuronyl-D-galactosyl-D-galactosyl-D-xylosyl-L-serine; heme P450-bis-L-cysteine-L-lysine; hexakis-L-cysteinyl hexairon hexasulfide; keratan sulfate D-glucuronyl-D-galactosyl-D-galactosyl-D-xylosyl-L-threonine; L-15 oxoalanine- lactic acid; L phenyllactic acid; l'-(8alpha-FAD)-L-histidine; L-2'.4',5'-topaquinone; L-3',4'-dihydroxyphenylalanine; L-3'.4'.5'-trihydroxyphenylalanine; L-4'-bromophenylalanine; L-6'-bromotryptophan; L-alanine amide; L-alanyl imidazolinone glycine; L-allysine; L-arginine amide; L-asparagine amide; L-aspartic 4-phosphoric anhydride; L-aspartic acid 1-amide; L-beta-methylthioaspartic acid; L-bromohistidine; L-20 citrulline; L-cysteine amide; L-cysteine glutathione disulfide; L-cysteine methyl disulfide; L-cysteine methyl ester; L-cysteine oxazolecarboxylic acid; L-cysteine oxazolinecarboxylic acid; L-cysteine persulfide; L-cysteine sulfenic acid; L-cysteine sulfinic acid; L-cysteine thiazolecarboxylic acid; L-cysteinyl homocitryl molybdenum-heptairon-nonasulfide; L-cysteinyl imidazolinone glycine; L-cysteinyl molybdopterin; L-25 cysteinyl molybdopterin guanine dinucleotide; L-cystine; L-erythro-beta-hydroxyasparagine; L-erythro-beta-hydroxyaspartic acid; L-gamma-carboxyglutamic acid; L-glutamic acid 1-amide; L-glutamic acid 5-methyl ester; L-glutamine amide; L-glutamyl 5-glycerolphosphorylethanolamine; L-histidine amide; L-isoglutamyl-polyglutamic acid; L-isoglutamyl-polyglycine; L-isoleucine amide; L-lanthionine; L-leucine amide; L-lysine 30 amide; L-lysine thiazolecarboxylic acid; L-lysinoalanine; L-methionine amide; L-methionine sulfone; L-phenylalanine thiazolecarboxylic acid; L-phenylalanine amide; L-proline amide; L-selenocysteine; L-selenocysteinyl molybdopterin guanine dinucleotide; L-serine amide; L-serine thiazolecarboxylic acid; L-seryl imidazolinone glycine; L-T-

- bromophenylalanine; L-T-bromophenylalanine; L-threonine amide; L-thyroxine; L-tryptophan amide; L-tryptophyl quinone; L-tyrosine amide; L-valine amide; meso-lanthionine; N-(L-glutamyl)-L-tyrosine; N-(L-isoaspartyl)-glycine; N-(L-isoaspartyl)-L-cysteine; N,N,N-trimethyl-L-alanine; N,N-dimethyl-L-proline; N2-acetyl-L-lysine; N2-succinyl-L-tryptophan; N4-(ADP-ribosyl)-L-asparagine; N4-glycosyl-L-asparagine; N4-hydroxymethyl-L-asparagine; N4-methyl-L-asparagine; N5-methyl-L-glutamine; N6-1-carboxyethyl-L-lysine; N6-(4-amino hydroxybutyl)-L-lysine; N6-(L-isoglutamyl)-L-lysine; N6-(phospho-5'-adenosine)-L-lysine; N6-(phospho-5'-guanosine)-L-lysine; N6,N6,N6-trimethyl-L-lysine; N6,N6-dimethyl-L-lysine; N6-acetyl-L-lysine; N6-biotinyl-L-lysine; N6-carboxy-L-lysine; N6-formyl-L-lysine; N6-glycyl-L-lysine; N6-lipoyl-L-lysine; N6-methyl-L-lysine; N6-methyl-N6-poly(N-methyl-propylamine)-L-lysine; N6-mureinyl-L-lysine; N6-myristoyl-L-lysine; N6-palmitoyl-L-lysine; N6-pyridoxal phosphate-L-lysine; N6-pyruvic acid 2-iminyl-L-lysine; N6-retinal-L-lysine; N-acetyl-glycine; N-acetyl-L-glutamine; N-acetyl-L-alanine; N-acetyl-L-aspartic acid; N-acetyl-L-cysteine; N-acetyl-L-glutamic acid; N-acetyl-L-isoleucine; N-acetyl-L-methionine; N-acetyl-L-proline; N-acetyl-L-serine; N-acetyl-L-threonine; N-acetyl-L-tyrosine; N-acetyl-L-valine; N-alanyl-glycosylphosphatidylinositoethanolamine; N-asparaginyl-glycosylphosphatidylinositoethanolamine; N-aspartyl-glycosylphosphatidylinositoethanolamine; N-formylglycine; N-formyl-L-methionine; N-glycyl-glycosylphosphatidylinositoethanolamine; N-L-glutamyl-poly-L-glutamic acid; N-methylglycine; N-methyl-L-alanine; N-methyl-L-methionine; N-methyl-L-phenylalanine; N-myristoyl-glycine; N-palmitoyl-L-cysteine; N-pyruvic acid 2-iminyl-L-cysteine; N-pyruvic acid 2-iminyl-L-valine; N-seryl-glycosylphosphatidylinositoethanolamine; N-seryl-glycosylsphingolipidinositoethanolamine; O-(ADP-ribosyl)-L-serine; O-(phospho-5'-adenosine)-L-threonine; O-(phospho-5'-DNA)-L-serine; O-(phospho-5'-DNA)-L-threonine; O-(phospho-5'rRNA)-L-serine; O-(phosphoribosyl dephospho-coenzyme A)-L-serine; O-(sn-1-glycerophosphoryl)-L-serine; O4'-(8alpha-FAD)-L-tyrosine; O4'-(phospho-5'-adenosine)-L-tyrosine; O4'-(phospho-5'-DNA)-L-tyrosine; O4'-(phospho-5'-RNA)-L-tyrosine; O4'-(phospho-5'-uridine)-L-tyrosine; O4-glycosyl-L-hydroxyproline; O4'-glycosyl-L-tyrosine; O4'-sulfo-L-tyrosine; O5-glycosyl-L-hydroxylysine; O-glycosyl-L-serine; O-glycosyl-L-threonine; omega-N-(ADP-ribosyl)-L-arginine; omega-N,omega-N'-dimethyl-L-arginine; omega-N-methyl-L-arginine; omega-N,omega-N'-dimethyl-L-arginine; omega-N-phospho-L-arginine; O-octanoyl-L-serine; O-palmitoyl-L-serine; O-

palmitoyl-L-threonine; O-phospho-L-serine; O-phospho-L-threonine; O-phosphopantetheine-L-serine; phycoerythrobilin-bis-L-cysteine; phycourobilin-bis-L-cysteine; pyrroloquinoline quinone; pyruvic acid; S hydroxycinnamyl-L-cysteine; S-(2-aminovinyl) methyl-D-cysteine; S-(2-aminovinyl)-D-cysteine; S-(6-FW-L-cysteine; S-
 5 (8 α -FAD)-L-cysteine; S-(ADP-ribosyl)-L-cysteine; S-(L-isoglutamyl)-L-cysteine; S-12-hydroxyfarnesyl-L-cysteine; S-acetyl-L-cysteine; S-diacylglycerol-L-cysteine; S-diphytanylglycerol diether-L-cysteine; S-farnesyl-L-cysteine; S-geranylgeranyl-L-cysteine; S-glycosyl-L-cysteine; S-glycyl-L-cysteine; S-methyl-L-cysteine; S-nitrosyl-L-cysteine; S-palmitoyl-L-cysteine; S-phospho-L-cysteine; S-phycobiliviolin-L-cysteine; S-
 10 phycocyanobilin-L-cysteine; S-phycoerythrobilin-L-cysteine; S-phytochromobilin-L-cysteine; S-selenyl-L-cysteine; S-sulfo-L-cysteine; tetrakis-L-cysteinyl diiron disulfide; tetrakis-L-cysteinyl iron; tetrakis-L-cysteinyl tetrairon tetrasulfide; trans-2,3-cis 4-dihydroxy-L-proline; tris-L-cysteinyl triiron tetrasulfide; tris-L-cysteinyl triiron trisulfide; tris-L-cysteinyl-L-aspartato tetrairon tetrasulfide; tris-L-cysteinyl-L-cysteine persulfido-
 15 bis-L-glutamato-L-histidino tetrairon disulfide trioxide; tris-L-cysteinyl-L-N^{3'}-histidino tetrairon tetrasulfide; tris-L-cysteinyl-L-N^{1'}-histidino tetrairon tetrasulfide; and tris-L-cysteinyl-L-serinyl tetrairon tetrasulfide.

Additional examples of PTMs may be found in web sites such as the Delta Mass database based on Krishna, R. G. and F. Wold (1998). Posttranslational Modifications.
 20 Proteins - Analysis and Design. R. H. Angeletti. San Diego, Academic Press. 1: 121-206. ; Methods in Enzymology, 193, J.A. McClosky (ed) (1990), pages 647-660; Methods in Protein Sequence Analysis edited by Kazutomo Imahori and Fumio Sakiyama, Plenum Press, (1993) "Post-translational modifications of proteins" R.G. Krishna and F. Wold pages 167-172; "GlycoSuiteDB: a new curated relational database of glycoprotein glycan
 25 structures and their biological sources" Cooper et al. Nucleic Acids Res. 29; 332-335 (2001) "O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins" Gupta et al. Nucleic Acids Research, 27: 370-372 (1999); and "PhosphoBase, a database of phosphorylation sites: release 2.0.", Kreegipuu et al. Nucleic Acids Res 27(1):237-239 (1999) see also, WO 02/21139A2.

30 Tumorigenesis is often accompanied by alterations in the post-translational modifications of proteins. Thus, in another embodiment, the invention provides polypeptides from cancerous cells or tissues that have altered post-translational modifications compared to the post-translational modifications of polypeptides from

normal cells or tissues. A number of altered post-translational modifications are known. One common alteration is a change in phosphorylation state, wherein the polypeptide from the cancerous cell or tissue is hyperphosphorylated or hypophosphorylated compared to the polypeptide from a normal tissue, or wherein the polypeptide is phosphorylated on
5 different residues than the polypeptide from a normal cell. Another common alteration is a change in glycosylation state, wherein the polypeptide from the cancerous cell or tissue has more or less glycosylation than the polypeptide from a normal tissue, and/or wherein the polypeptide from the cancerous cell or tissue has a different type of glycosylation than the polypeptide from a noncancerous cell or tissue. Changes in glycosylation may be
10 critical because carbohydrate-protein and carbohydrate-carbohydrate interactions are important in cancer cell progression, dissemination and invasion. See, e.g., Barchi, Curr. Pharm. Des. 6: 485-501 (2000), Verma, Cancer Biochem. Biophys. 14: 151-162 (1994) and Dennis et al., Bioessays 5: 412-421 (1999).

Another post-translational modification that may be altered in cancer cells is
15 prenylation. Prenylation is the covalent attachment of a hydrophobic prenyl group (either farnesyl or geranylgeranyl) to a polypeptide. Prenylation is required for localizing a protein to a cell membrane and is often required for polypeptide function. For instance, the Ras superfamily of GTPase signalling proteins must be prenylated for function in a cell. See, e.g., Prendergast et al., Semin. Cancer Biol. 10: 443-452 (2000) and Khwaja et al., Lancet 355: 741-744 (2000).
20

Other post-translation modifications that may be altered in cancer cells include, without limitation, polypeptide methylation, acetylation, arginylation or racemization of amino acid residues. In these cases, the polypeptide from the cancerous cell may exhibit either increased or decreased amounts of the post-translational modification compared to
25 the corresponding polypeptides from noncancerous cells.

Other polypeptide alterations in cancer cells include abnormal polypeptide cleavage of proteins and aberrant protein-protein interactions. Abnormal polypeptide cleavage may be cleavage of a polypeptide in a cancerous cell that does not usually occur in a normal cell, or a lack of cleavage in a cancerous cell, wherein the polypeptide is
30 cleaved in a normal cell. Aberrant protein-protein interactions may be either covalent cross-linking or non-covalent binding between proteins that do not normally bind to each other. Alternatively, in a cancerous cell, a protein may fail to bind to another protein to which it is bound in a noncancerous cell. Alterations in cleavage or in protein-protein

interactions may be due to over- or underproduction of a polypeptide in a cancerous cell compared to that in a normal cell, or may be due to alterations in post-translational modifications (see above) of one or more proteins in the cancerous cell. See, e.g., Henschen-Edman, Ann. N.Y. Acad. Sci. 936: 580-593 (2001).

5 Alterations in polypeptide post-translational modifications, as well as changes in polypeptide cleavage and protein-protein interactions, may be determined by any method known in the art. For instance, alterations in phosphorylation may be determined by using anti-phosphoserine, anti-phosphothreonine or anti-phosphotyrosine antibodies or by amino acid analysis. Glycosylation alterations may be determined using antibodies specific for
10 different sugar residues, by carbohydrate sequencing, or by alterations in the size of the glycoprotein, which can be determined by, e.g., SDS polyacrylamide gel electrophoresis (PAGE). Other alterations of post-translational modifications, such as prenylation, racemization, methylation, acetylation and arginylation, may be determined by chemical analysis, protein sequencing, amino acid analysis, or by using antibodies specific for the
15 particular post-translational modifications. Changes in protein-protein interactions and in polypeptide cleavage may be analyzed by any method known in the art including, without limitation, non-denaturing PAGE (for non-covalent protein-protein interactions), SDS PAGE (for covalent protein-protein interactions and protein cleavage), chemical cleavage, protein sequencing or immunoassays.

20 In another embodiment, the invention provides polypeptides that have been post-translationally modified. In one embodiment, polypeptides may be modified enzymatically or chemically, by addition or removal of a post-translational modification. For example, a polypeptide may be glycosylated or deglycosylated enzymatically. Similarly, polypeptides may be phosphorylated using a purified kinase, such as a MAP
25 kinase (e.g., p38, ERK, or JNK) or a tyrosine kinase (e.g., Src or erbB2). A polypeptide may also be modified through synthetic chemistry. Alternatively, one may isolate the polypeptide of interest from a cell or tissue that expresses the polypeptide with the desired post-translational modification. In another embodiment, a nucleic acid molecule encoding the polypeptide of interest is introduced into a host cell that is capable of post-
30 translationally modifying the encoded polypeptide in the desired fashion. If the polypeptide does not contain a motif for a desired post-translational modification, one may alter the post-translational modification by mutating the nucleic acid sequence of a nucleic acid molecule encoding the polypeptide so that it contains a site for the desired post-

translational modification. Amino acid sequences that may be post-translationally modified are known in the art. See, e.g., the programs described above on the website www.expasy.org. The nucleic acid molecule may also be introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide. Similarly, one may delete sites that are post-translationally modified by either mutating the nucleic acid sequence so that the encoded polypeptide does not contain the post-translational modification motif, or by introducing the native nucleic acid molecule into a host cell that is not capable of post-translationally modifying the encoded polypeptide.

It will be appreciated, as is well-known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

Useful post-synthetic (and post-translational) modifications include conjugation to detectable labels, such as fluorophores. A wide variety of amine-reactive and thiol-reactive fluorophore derivatives have been synthesized that react under nondenaturing conditions with N-terminal amino groups and epsilon amino groups of lysine residues, on the one hand, and with free thiol groups of cysteine residues, on the other.

Kits are available commercially that permit conjugation of proteins to a variety of amine-reactive or thiol-reactive fluorophores: Molecular Probes, Inc. (Eugene, OR, USA), e.g., offers kits for conjugating proteins to Alexa Fluor 350, Alexa Fluor 430, Fluorescein-EX, Alexa Fluor 488, Oregon Green 488, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 546, Alexa Fluor 568, Alexa Fluor 594, and Texas Red-X.

A wide variety of other amine-reactive and thiol-reactive fluorophores are available commercially (Molecular Probes, Inc., Eugene, OR, USA), including Alexa

Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568,
 5 BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA).

10 The polypeptides of the present invention can also be conjugated to fluorophores, other proteins, and other macromolecules, using bifunctional linking reagents. Common homobifunctional reagents include, *e.g.*, APG, AEDP, BASED, BMB, BMDB, BMH, BMOE, BM[PEO]3, BM[PEO]4, BS3, BSOCEs, DFDNB, DMA, DMP, DMS, DPDPB, DSG, DSP (Lomant's Reagent), DSS, DST, DTBP, DTME, DTSSP, EGS, HBVS,
 15 Sulfo-BSOCEs, Sulfo-DST, Sulfo-EGS (all available from Pierce, Rockford, IL, USA); common heterobifunctional cross-linkers include ABH, AMAS, ANB-NOS, APDP, ASBA, BMPEA, BMPH, BMPS, EDC, EMCA, EMCH, EMCS, KMUA, KMH, GMBS, LC-SMCC, LC-SPDP, MBS, M2C2H, MPBH, MSA, NHS-ASA, PDPH, PMPI, SADP, SAED, SAND, SANPAH, SASD, SATP, SBAP, SFAD, SIA, SIAB, SMCC, SMPB,
 20 SMPH, SMPT, SPDP, Sulfo-EMCS, Sulfo-GMBS, Sulfo-HSAB, Sulfo-KMUS, Sulfo-LC-SPDP, Sulfo-MBS, Sulfo-NHS-LC-ASA, Sulfo-SADP, Sulfo-SANPAH, Sulfo-SIAB, Sulfo-SMCC, Sulfo-SMPB, Sulfo-LC-SMPT, SVSB, TFCS (all available Pierce, Rockford, IL, USA).

Polypeptides of the present invention, including full length polypeptides,
 25 fragments and fusion proteins, can be conjugated, using such cross-linking reagents, to fluorophores that are not amine- or thiol-reactive. Other labels that usefully can be conjugated to polypeptides of the present invention include radioactive labels, echosonographic contrast reagents, and MRI contrast agents.

Polypeptides of the present invention, including full length polypeptide, fragments
 30 and fusion proteins, can also usefully be conjugated using cross-linking agents to carrier proteins, such as KLH, bovine thyroglobulin, and even bovine serum albumin (BSA), to increase immunogenicity for raising anti-CSP antibodies.

Polypeptides of the present invention, including full length polypeptide, fragments and fusion proteins, can also usefully be conjugated to polyethylene glycol (PEG); PEGylation increases the serum half life of proteins administered intravenously for replacement therapy. Delgado *et al.*, *Crit. Rev. Ther. Drug Carrier Syst.* 9(3-4): 249-304 (1992); Scott *et al.*, *Curr. Pharm. Des.* 4(6): 423-38 (1998); DeSantis *et al.*, *Curr. Opin. Biotechnol.* 10(4): 324-30 (1999). PEG monomers can be attached to the protein directly or through a linker, with PEGylation using PEG monomers activated with tresyl chloride (2,2,2-trifluoroethanesulphonyl chloride) permitting direct attachment under mild conditions.

10 Polypeptides of the present invention are also inclusive of analogs of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, this polypeptide is a CSP. In a more preferred embodiment, this polypeptide is derived from a polypeptide having part or all of the amino acid sequence of SEQ ID NO: 101-190. Also preferred is an analog polypeptide comprising one or more
15 substitutions of non-natural amino acids or non-native inter-residue bonds compared to the naturally-occurring polypeptide. In one embodiment, the analog is structurally similar to a CSP, but one or more peptide linkages is replaced by a linkage selected from the group consisting of --CH₂NH--, --CH₂S--, --CH₂-CH₂--, --CH=CH--(cis and trans), --COCH₂--, --CH(OH)CH₂-- and --CH₂SO--. In another embodiment, the analog comprises
20 substitution of one or more amino acids of a CSP with a D-amino acid of the same type or other non-natural amino acid in order to generate more stable peptides. D-amino acids can readily be incorporated during chemical peptide synthesis: peptides assembled from D-amino acids are more resistant to proteolytic attack; incorporation of D-amino acids can also be used to confer specific three dimensional conformations on the peptide. Other
25 amino acid analogues commonly added during chemical synthesis include ornithine, norleucine, phosphorylated amino acids (typically phosphoserine, phosphothreonine, phosphotyrosine), L-malonyltyrosine, a non-hydrolyzable analog of phosphotyrosine (*see, e.g., Krole et al., Biochem. Biophys. Res. Com.* 209: 817-821 (1995)), and various halogenated phenylalanine derivatives.

30 Non-natural amino acids can be incorporated during solid phase chemical synthesis or by recombinant techniques, although the former is typically more common. Solid phase chemical synthesis of peptides is well established in the art. Procedures are described, *inter alia*, in Chan *et al.* (eds.), Fmoc Solid Phase Peptide Synthesis: A

Practical Approach (Practical Approach Series), Oxford Univ. Press (March 2000); Jones, Amino Acid and Peptide Synthesis (Oxford Chemistry Primers, No 7), Oxford Univ. Press (1992); and Bodanszky, Principles of Peptide Synthesis (Springer Laboratory), Springer Verlag (1993).

5 Amino acid analogues having detectable labels are also usefully incorporated during synthesis to provide derivatives and analogs. Biotin, for example can be added using biotinoyl--(9-fluorenylmethoxycarbonyl)-L-lysine (Fmoc biocytin) (Molecular Probes, Eugene, OR, USA). Biotin can also be added enzymatically by incorporation into a fusion protein of a *E. coli* BirA substrate peptide. The Fmoc and *t*BOC derivatives of
 10 dabcyL-L-lysine (Molecular Probes, Inc., Eugene, OR, USA) can be used to incorporate the dabcyL chromophore at selected sites in the peptide sequence during synthesis. The aminonaphthalene derivative EDANS, the most common fluorophore for pairing with the dabcyL quencher in fluorescence resonance energy transfer (FRET) systems, can be introduced during automated synthesis of peptides by using EDANS--Fmoc-L-glutamic
 15 acid or the corresponding *t*BOC derivative (both from Molecular Probes, Inc., Eugene, OR, USA). Tetramethylrhodamine fluorophores can be incorporated during automated Fmoc synthesis of peptides using (Fmoc)--TMR-L-lysine (Molecular Probes, Inc. Eugene, OR, USA).

Other useful amino acid analogues that can be incorporated during chemical
 20 synthesis include aspartic acid, glutamic acid, lysine, and tyrosine analogues having allyl side-chain protection (Applied Biosystems, Inc., Foster City, CA, USA); the allyl side chain permits synthesis of cyclic, branched-chain, sulfonated, glycosylated, and phosphorylated peptides.

A large number of other Fmoc-protected non-natural amino acid analogues
 25 capable of incorporation during chemical synthesis are available commercially, including, e.g., Fmoc-2-aminobicyclo[2.2.1]heptane-2-carboxylic acid, Fmoc-3-endo-aminobicyclo[2.2.1]heptane-2-endo-carboxylic acid, Fmoc-3-exo-aminobicyclo[2.2.1]heptane-2-exo-carboxylic acid, Fmoc-3-endo-amino-bicyclo[2.2.1]hept-5-ene-2-endo-carboxylic acid, Fmoc-3-exo-amino-bicyclo[2.2.1]hept-
 30 5-ene-2-exo-carboxylic acid, Fmoc-cis-2-amino-1-cyclohexanecarboxylic acid, Fmoc-trans-2-amino-1-cyclohexanecarboxylic acid, Fmoc-1-amino-1-cyclopentanecarboxylic acid, Fmoc-cis-2-amino-1-cyclopentanecarboxylic acid, Fmoc-1-amino-1-cyclopropanecarboxylic acid, Fmoc-D-2-amino-4-(ethylthio)butyric acid, Fmoc-L-2-

amino-4-(ethylthio)butyric acid, Fmoc-L-buthionine, Fmoc-S-methyl-L-Cysteine, Fmoc-2-aminobenzoic acid (anthranillic acid), Fmoc-3-aminobenzoic acid, Fmoc-4-aminobenzoic acid, Fmoc-2-aminobenzophenone-2'-carboxylic acid, Fmoc-N-(4-aminobenzoyl)- β -alanine, Fmoc-2-amino-4,5-dimethoxybenzoic acid, Fmoc-4-aminohippuric acid, Fmoc-2-amino-3-hydroxybenzoic acid, Fmoc-2-amino-5-hydroxybenzoic acid, Fmoc-3-amino-4-hydroxybenzoic acid, Fmoc-4-amino-3-hydroxybenzoic acid, Fmoc-4-amino-2-hydroxybenzoic acid, Fmoc-5-amino-2-hydroxybenzoic acid, Fmoc-2-amino-3-methoxybenzoic acid, Fmoc-4-amino-3-methoxybenzoic acid, Fmoc-2-amino-3-methylbenzoic acid, Fmoc-2-amino-5-methylbenzoic acid, Fmoc-2-amino-6-methylbenzoic acid, Fmoc-3-amino-2-methylbenzoic acid, Fmoc-3-amino-4-methylbenzoic acid, Fmoc-4-amino-3-methylbenzoic acid, Fmoc-3-amino-2-naphtoic acid, Fmoc-D,L-3-amino-3-phenylpropionic acid, Fmoc-L-Methyldopa, Fmoc-2-amino-4,6-dimethyl-3-pyridinecarboxylic acid, Fmoc-D,L-amino-2-thiophenacetic acid, Fmoc-4-(carboxymethyl)piperazine, Fmoc-4-carboxypiperazine, Fmoc-4-(carboxymethyl)homopiperazine, Fmoc-4-phenyl-4-piperidinecarboxylic acid, Fmoc-L-1,2,3,4-tetrahydronorharman-3-carboxylic acid, Fmoc-L-thiazolidine-4-carboxylic acid, all available from The Peptide Laboratory (Richmond, CA, USA).

Non-natural residues can also be added biosynthetically by engineering a suppressor tRNA, typically one that recognizes the UAG stop codon, by chemical aminoacylation with the desired unnatural amino acid. Conventional site-directed mutagenesis is used to introduce the chosen stop codon UAG at the site of interest in the protein gene. When the acylated suppressor tRNA and the mutant gene are combined in an *in vitro* transcription/translation system, the unnatural amino acid is incorporated in response to the UAG codon to give a protein containing that amino acid at the specified position. Liu *et al.*, *Proc. Natl Acad. Sci. USA* 96(9): 4780-5 (1999); Wang *et al.*, *Science* 292(5516): 498-500 (2001).

Fusion Proteins

Another aspect of the present invention relates to the fusion of a polypeptide of the present invention to heterologous polypeptides. In a preferred embodiment, the polypeptide of the present invention is a CSP. In a more preferred embodiment, the polypeptide of the present invention that is fused to a heterologous polypeptide comprises part or all of the amino acid sequence of SEQ ID NO: 101-190, or is a mutein,

homologous polypeptide, analog or derivative thereof. In an even more preferred embodiment, the fusion protein is encoded by a nucleic acid molecule comprising all or part of the nucleic acid sequence of SEQ ID NO: 1-100, or comprises all or part of a nucleic acid sequence that selectively hybridizes or is homologous to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1-100.

The fusion proteins of the present invention will include at least one fragment of a polypeptide of the present invention, which fragment is at least 6, typically at least 8, often at least 15, and usefully at least 16, 17, 18, 19, or 20 amino acids long. The fragment of the polypeptide of the present to be included in the fusion can usefully be at least 25 amino acids long, at least 50 amino acids long, and can be at least 75, 100, or even 150 amino acids long. Fusions that include the entirety of a polypeptide of the present invention have particular utility.

The heterologous polypeptide included within the fusion protein of the present invention is at least 6 amino acids in length, often at least 8 amino acids in length, and preferably at least 15, 20, or 25 amino acids in length. Fusions that include larger polypeptides, such as the IgG Fc region, and even entire proteins (such as GFP chromophore-containing proteins) are particularly useful.

As described above in the description of vectors and expression vectors of the present invention, which discussion is incorporated here by reference in its entirety, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those designed to facilitate purification and/or visualization of recombinantly-expressed proteins. *See, e.g., Ausubel, Chapter 16, (1992), supra.* Although purification tags can also be incorporated into fusions that are chemically synthesized, chemical synthesis typically provides sufficient purity that further purification by HPLC suffices; however, visualization tags as above described retain their utility even when the protein is produced by chemical synthesis, and when so included render the fusion proteins of the present invention useful as directly detectable markers of the presence of a polypeptide of the invention.

As also discussed above, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those that facilitate secretion of recombinantly expressed proteins into the periplasmic space or extracellular milieu for prokaryotic hosts or into the culture medium for eukaryotic cells through incorporation of secretion signals and/or leader sequences. For example, a His⁶ tagged protein can be

purified on a Ni affinity column and a GST fusion protein can be purified on a glutathione affinity column. Similarly, a fusion protein comprising the Fc domain of IgG can be purified on a Protein A or Protein G column and a fusion protein comprising an epitope tag such as myc can be purified using an immunoaffinity column containing an anti-c-myc antibody. It is preferable that the epitope tag be separated from the protein encoded by the essential gene by an enzymatic cleavage site that can be cleaved after purification. See also the discussion of nucleic acid molecules encoding fusion proteins that may be expressed on the surface of a cell.

Other useful fusion proteins of the present invention include those that permit use of the polypeptide of the present invention as bait in a yeast two-hybrid system. See Bartel *et al.* (eds.), *The Yeast Two-Hybrid System*, Oxford University Press (1997); Zhu *et al.*, *Yeast Hybrid Technologies*, Eaton Publishing (2000); Fields *et al.*, *Trends Genet.* 10(8): 286-92 (1994); Mendelsohn *et al.*, *Curr. Opin. Biotechnol.* 5(5): 482-6 (1994); Luban *et al.*, *Curr. Opin. Biotechnol.* 6(1): 59-64 (1995); Allen *et al.*, *Trends Biochem. Sci.* 20(12): 511-6 (1995); Drees, *Curr. Opin. Chem. Biol.* 3(1): 64-70 (1999); Topcu *et al.*, *Pharm. Res.* 17(9): 1049-55 (2000); Fashena *et al.*, *Gene* 250(1-2): 1-14 (2000); ; Colas *et al.*, (1996) Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* 380, 548-550; Norman, T. *et al.*, (1999) Genetic selection of peptide inhibitors of biological pathways. *Science* 285, 591-595, Fabbri *et al.*, (1999) Inhibition of mammalian cell proliferation by genetically selected peptide aptamers that functionally antagonize E2F activity. *Oncogene* 18, 4357-4363; Xu *et al.*, (1997) Cells that register logical relationships among proteins. *Proc Natl Acad Sci U S A* 94, 12473-12478; Yang, *et al.*, (1995) Protein-peptide interactions analyzed with the yeast two-hybrid system. *Nuc. Acids Res.* 23, 1152-1156; Kolonin *et al.*, (1998) Targeting cyclin-dependent kinases in *Drosophila* with peptide aptamers. *Proc Natl Acad Sci U S A* 95, 14266-14271; Cohen *et al.*, (1998) An artificial cell-cycle inhibitor isolated from a combinatorial library. *Proc Natl Acad Sci U S A* 95, 14272-14277; Uetz, P.; Giot, L.; al, e.; Fields, S.; Rothberg, J. M. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403, 623-627; Ito, *et al.*, (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 98, 4569-4574. Typically, such fusion is to either *E. coli* LexA or yeast GAL4 DNA binding domains. Related bait plasmids are available that express the bait fused to a nuclear localization signal.

Other useful fusion proteins include those that permit display of the encoded polypeptide on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as green fluorescent protein (GFP), and fusions to the IgG Fc region, as described above, which discussion is incorporated here by reference in its entirety.

5 The polypeptides of the present invention can also usefully be fused to protein toxins, such as *Pseudomonas* exotoxin A, diphtheria toxin, shiga toxin A, anthrax toxin lethal factor, ricin, in order to effect ablation of cells that bind or take up the proteins of the present invention.

10 Fusion partners include, *inter alia*, *myc*, hemagglutinin (HA), GST, immunoglobulins, β -galactosidase, biotin trpE, protein A, β -lactamase, α -amylase, maltose binding protein, alcohol dehydrogenase, polyhistidine (for example, six histidine at the amino and/or carboxyl terminus of the polypeptide), lacZ, green fluorescent protein (GFP), yeast α mating factor, GAL4 transcription activation or DNA binding domain, luciferase, and serum proteins such as ovalbumin, albumin and the constant domain of
15 IgG. *See, e.g.*, Ausubel (1992), *supra* and Ausubel (1999), *supra*. Fusion proteins may also contain sites for specific enzymatic cleavage, such as a site that is recognized by enzymes such as Factor XIII, trypsin, pepsin, or any other enzyme known in the art. Fusion proteins will typically be made by either recombinant nucleic acid methods, as described above, chemically synthesized using techniques well-known in the art (*e.g.*, a
20 Merrifield synthesis), or produced by chemical cross-linking.

Another advantage of fusion proteins is that the epitope tag can be used to bind the fusion protein to a plate or column through an affinity linkage for screening binding proteins or other molecules that bind to the CSP.

25 As further described below, the polypeptides of the present invention can readily be used as specific immunogens to raise antibodies that specifically recognize polypeptides of the present invention including CSPs and their allelic variants and homologues. The antibodies, in turn, can be used, *inter alia*, specifically to assay for the polypeptides of the present invention, particularly CSPs, *e.g.* by ELISA for detection of protein fluid samples, such as serum, by immunohistochemistry or laser scanning
30 cytometry, for detection of protein in tissue samples, or by flow cytometry, for detection of intracellular protein in cell suspensions, for specific antibody-mediated isolation and/or purification of CSPs, as for example by immunoprecipitation, and for use as specific agonists or antagonists of CSPs.

One may determine whether polypeptides of the present invention including CSPs, muteins, homologous proteins or allelic variants or fusion proteins of the present invention are functional by methods known in the art. For instance, residues that are tolerant of change while retaining function can be identified by altering the polypeptide at known
5 residues using methods known in the art, such as alanine scanning mutagenesis, Cunningham *et al.*, *Science* 244(4908): 1081-5 (1989); transposon linker scanning mutagenesis, Chen *et al.*, *Gene* 263(1-2): 39-48 (2001); combinations of homolog- and alanine-scanning mutagenesis, Jin *et al.*, *J. Mol. Biol.* 226(3): 851-65 (1992); combinatorial alanine scanning, Weiss *et al.*, *Proc. Natl. Acad. Sci USA* 97(16): 8950-4
10 (2000), followed by functional assay. Transposon linker scanning kits are available commercially (New England Biolabs, Beverly, MA, USA, catalog. no. E7-102S; EZ::TN™ In-Frame Linker Insertion Kit, catalogue no. EZI04KN, Epicentre Technologies Corporation, Madison, WI, USA).

Purification of the polypeptides or fusion proteins of the present invention is well-
15 known and within the skill of one having ordinary skill in the art. *See, e.g.*, Scopes, Protein Purification, 2d ed. (1987). Purification of recombinantly expressed polypeptides is described above. Purification of chemically-synthesized peptides can readily be effected, *e.g.*, by HPLC.

Accordingly, it is an aspect of the present invention to provide the isolated
20 polypeptides or fusion proteins of the present invention in pure or substantially pure form in the presence of absence of a stabilizing agent. Stabilizing agents include both proteinaceous or non-proteinaceous material and are well-known in the art. Stabilizing agents, such as albumin and polyethylene glycol (PEG) are known and are commercially available.

25 Although high levels of purity are preferred when the isolated polypeptide or fusion protein of the present invention are used as therapeutic agents, such as in vaccines and replacement therapy, the isolated polypeptides of the present invention are also useful at lower purity. For example, partially purified polypeptides of the present invention can be used as immunogens to raise antibodies in laboratory animals.

30 In a preferred embodiment, the purified and substantially purified polypeptides of the present invention are in compositions that lack detectable ampholytes, acrylamide monomers, bis-acrylamide monomers, and polyacrylamide.

The polypeptides or fusion proteins of the present invention can usefully be attached to a substrate. The substrate can be porous or solid, planar or non-planar; the bond can be covalent or noncovalent.

For example, the polypeptides or fusion proteins of the present invention can usefully be bound to a porous substrate, commonly a membrane, typically comprising nitrocellulose, polyvinylidene fluoride (PVDF), or cationically derivatized, hydrophilic PVDF; so bound, the polypeptides or fusion proteins of the present invention can be used to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized polypeptide or fusion protein of the present invention.

As another example, the polypeptides or fusion proteins of the present invention can usefully be bound to a substantially nonporous substrate, such as plastic, to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention. Such plastics include polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof; when the assay is performed in a standard microtiter dish, the plastic is typically polystyrene.

The polypeptides and fusion proteins of the present invention can also be attached to a substrate suitable for use as a surface enhanced laser desorption ionization source; so attached, the polypeptide or fusion protein of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound polypeptide or fusion protein to indicate biologic interaction there between. The polypeptides or fusion proteins of the present invention can also be attached to a substrate suitable for use in surface plasmon resonance detection; so attached, the polypeptide or fusion protein of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound polypeptide or fusion protein to indicate biological interaction there between.

Antibodies

In another aspect, the invention provides antibodies, including fragments and derivatives thereof, that bind specifically to polypeptides encoded by the nucleic acid molecules of the invention. In a preferred embodiment, the antibodies are specific for a polypeptide that is a CSP, or a fragment, mutein, derivative, analog or fusion protein

thereof. In a more preferred embodiment, the antibodies are specific for a polypeptide that comprises SEQ ID NO: 101-190, or a fragment, mutein, derivative, analog or fusion protein thereof.

The antibodies of the present invention can be specific for linear epitopes,
5 discontinuous epitopes, or conformational epitopes of such proteins or protein fragments, either as present on the protein in its native conformation or, in some cases, as present on the proteins as denatured, as, *e.g.*, by solubilization in SDS. New epitopes may be also due to a difference in post translational modifications (PTMs) in disease versus normal tissue. For example, a particular site on a CSP may be glycosylated in cancerous cells, but
10 not glycosylated in normal cells or vis versa. In addition, alternative splice forms of a CSP may be indicative of cancer. Differential degradation of the C or N-terminus of a CSP may also be a marker or target for anticancer therapy. For example, an CSP may be N-terminal degraded in cancer cells exposing new epitopes to which antibodies may selectively bind for diagnostic or therapeutic uses.

15 As is well-known in the art, the degree to which an antibody can discriminate as among molecular species in a mixture will depend, in part, upon the conformational relatedness of the species in the mixture; typically, the antibodies of the present invention will discriminate over adventitious binding to non-CSP polypeptides by at least two-fold, more typically by at least 5-fold, typically by more than 10-fold, 25-fold, 50-fold, 75-fold,
20 and often by more than 100-fold, and on occasion by more than 500-fold or 1000-fold. When used to detect the proteins or protein fragments of the present invention, the antibody of the present invention is sufficiently specific when it can be used to determine the presence of the polypeptide of the present invention in samples derived from human colon.

25 Typically, the affinity or avidity of an antibody (or antibody multimer, as in the case of an IgM pentamer) of the present invention for a protein or protein fragment of the present invention will be at least about 1×10^{-6} molar (M), typically at least about 5×10^{-7} M, 1×10^{-7} M, with affinities and avidities of at least 1×10^{-8} M, 5×10^{-9} M, 1×10^{-10} M and up to 1×10^{-13} M proving especially useful.

30 The antibodies of the present invention can be naturally-occurring forms, such as IgG, IgM, IgD, IgE, IgY, and IgA, from any avian, reptilian, or mammalian species.

Human antibodies can, but will infrequently, be drawn directly from human donors or human cells. In such case, antibodies to the polypeptides of the present invention will

typically have resulted from fortuitous immunization, such as autoimmune immunization, with the polypeptide of the present invention. Such antibodies will typically, but will not invariably, be polyclonal. In addition, individual polyclonal antibodies may be isolated and cloned to generate monoclonals.

5 Human antibodies are more frequently obtained using transgenic animals that express human immunoglobulin genes, which transgenic animals can be affirmatively immunized with the protein immunogen of the present invention. Human Ig-transgenic mice capable of producing human antibodies and methods of producing human antibodies therefrom upon specific immunization are described, *inter alia*, in United States Patent
10 Nos. 6,162,963; 6,150,584; 6,114,598; 6,075,181; 5,939,598; 5,877,397; 5,874,299; 5,814,318; 5,789,650; 5,770,429; 5,661,016; 5,633,425; 5,625,126; 5,569,825; 5,545,807; 5,545,806, and 5,591,669, the disclosures of which are incorporated herein by reference in their entireties. Such antibodies are typically monoclonal, and are typically produced using techniques developed for production of murine antibodies.

15 Human antibodies are particularly useful, and often preferred, when the antibodies of the present invention are to be administered to human beings as *in vivo* diagnostic or therapeutic agents, since recipient immune response to the administered antibody will often be substantially less than that occasioned by administration of an antibody derived from another species, such as mouse.

20 IgG, IgM, IgD, IgE, IgY, and IgA antibodies of the present invention are also usefully obtained from other species, including mammals such as rodents (typically mouse, but also rat, guinea pig, and hamster), lagomorphs, (typically rabbits), and also larger mammals, such as sheep, goats, cows, and horses; or egg laying birds or reptiles such as chickens or alligators. In such cases, as with the transgenic human-antibody-
25 producing non-human mammals, fortuitous immunization is not required, and the non-human mammal is typically affirmatively immunized, according to standard immunization protocols, with the polypeptide of the present invention. One form of avian antibodies may be generated using techniques described in WO 00/29444, published 25 May 2000.

As discussed above, virtually all fragments of 8 or more contiguous amino acids of
30 a polypeptide of the present invention can be used effectively as immunogens when conjugated to a carrier, typically a protein such as bovine thyroglobulin, keyhole limpet hemocyanin, or bovine serum albumin, conveniently using a bifunctional linker such as those described elsewhere above, which discussion is incorporated by reference here.

Immunogenicity can also be conferred by fusion of the polypeptide of the present invention to other moieties. For example, polypeptides of the present invention can be produced by solid phase synthesis on a branched polylysine core matrix; these multiple antigenic peptides (MAPs) provide high purity, increased avidity, accurate chemical
5 definition and improved safety in vaccine development. Tam *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5409-5413 (1988); Posnett *et al.*, *J. Biol. Chem.* 263: 1719-1725 (1988).

Protocols for immunizing non-human mammals or avian species are well-established in the art. See Harlow *et al.* (eds.), Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory (1998); Coligan *et al.* (eds.), Current Protocols in
10 Immunology, John Wiley & Sons, Inc. (2001); Zola, Monoclonal Antibodies: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives (Basics: From Background to Bench), Springer Verlag (2000); Gross M, Speck *J.Dtsch. Tierarztl. Wochenschr.* 103: 417-422 (1996). Immunization protocols often include multiple immunizations, either with or without adjuvants such as Freund's complete adjuvant and
15 Freund's incomplete adjuvant, and may include naked DNA immunization (Moss, *Semin. Immunol.* 2: 317-327 (1990).

Antibodies from non-human mammals and avian species can be polyclonal or monoclonal, with polyclonal antibodies having certain advantages in immunohistochemical detection of the polypeptides of the present invention and
20 monoclonal antibodies having advantages in identifying and distinguishing particular epitopes of the polypeptides of the present invention. Antibodies from avian species may have particular advantage in detection of the polypeptides of the present invention, in human serum or tissues (Vikinge *et al.*, *Biosens. Bioelectron.* 13: 1257-1262 (1998). Following immunization, the antibodies of the present invention can be obtained using any
25 art-accepted technique. Such techniques are well-known in the art and are described in detail in references such as Coligan, *supra*; Zola, *supra*; Howard *et al.* (eds.), Basic Methods in Antibody Production and Characterization, CRC Press (2000); Harlow, *supra*; Davis (ed.), Monoclonal Antibody Protocols, Vol. 45, Humana Press (1995); Delves (ed.), Antibody Production: Essential Techniques, John Wiley & Son Ltd (1997); and Kenney,
30 Antibody Solution: An Antibody Methods Manual, Chapman & Hall (1997).

Briefly, such techniques include, *inter alia*, production of monoclonal antibodies by hybridomas and expression of antibodies or fragments or derivatives thereof from host cells engineered to express immunoglobulin genes or fragments thereof. These two

methods of production are not mutually exclusive: genes encoding antibodies specific for the polypeptides of the present invention can be cloned from hybridomas and thereafter expressed in other host cells. Nor need the two necessarily be performed together: *e.g.*, genes encoding antibodies specific for the polypeptides of the present invention can be
5 cloned directly from B cells known to be specific for the desired protein, as further described in United States Patent No. 5,627,052, the disclosure of which is incorporated herein by reference in its entirety, or from antibody-displaying phage.

Recombinant expression in host cells is particularly useful when fragments or derivatives of the antibodies of the present invention are desired.

10 Host cells for recombinant antibody production of whole antibodies, antibody fragments, or antibody derivatives can be prokaryotic or eukaryotic.

Prokaryotic hosts are particularly useful for producing phage displayed antibodies of the present invention.

The technology of phage-displayed antibodies, in which antibody variable region
15 fragments are fused, for example, to the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13, is by now well-established. *See, e.g.*, Sidhu, *Curr. Opin. Biotechnol.* 11(6): 610-6 (2000); Griffiths *et al.*, *Curr. Opin. Biotechnol.* 9(1): 102-8 (1998); Hoogenboom *et al.*, *Immunotechnology*, 4(1): 1-20 (1998); Rader *et al.*, *Current Opinion in Biotechnology* 8: 503-508 (1997); Aujame *et al.*, *Human*
20 *Antibodies* 8: 155-168 (1997); Hoogenboom, *Trends in Biotechnol.* 15: 62-70 (1997); de Kruif *et al.*, 17: 453-455 (1996); Barbas *et al.*, *Trends in Biotechnol.* 14: 230-234 (1996); Winter *et al.*, *Ann. Rev. Immunol.* 433-455 (1994). Techniques and protocols required to generate, propagate, screen (pan), and use the antibody fragments from such libraries have recently been compiled. *See, e.g.*, Barbas (2001), *supra*; Kay, *supra*; and Abelson, *supra*.

25 Typically, phage-displayed antibody fragments are scFv fragments or Fab fragments; when desired, full length antibodies can be produced by cloning the variable regions from the displaying phage into a complete antibody and expressing the full length antibody in a further prokaryotic or a eukaryotic host cell. Eukaryotic cells are also useful for expression of the antibodies, antibody fragments, and antibody derivatives of the
30 present invention. For example, antibody fragments of the present invention can be produced in *Pichia pastoris* and in *Saccharomyces cerevisiae*. *See, e.g.*, Takahashi *et al.*, *Biosci. Biotechnol. Biochem.* 64(10): 2138-44 (2000); Freyre *et al.*, *J. Biotechnol.* 76(2-3):1 57-63 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 117-20

(1999); Pennell *et al.*, *Res. Immunol.* 149(6): 599-603 (1998); Eldin *et al.*, *J. Immunol. Methods.* 201(1): 67-75 (1997);, Frenken *et al.*, *Res. Immunol.* 149(6): 589-99 (1998); and Shusta *et al.*, *Nature Biotechnol.* 16(8): 773-7 (1998).

Antibodies, including antibody fragments and derivatives, of the present invention
5 can also be produced in insect cells. See, e.g., Li *et al.*, *Protein Expr. Purif.* 21(1): 121-8 (2001); Ailor *et al.*, *Biotechnol. Bioeng.* 58(2-3): 196-203 (1998); Hsu *et al.*, *Biotechnol. Prog.* 13(1): 96-104 (1997); Edelman *et al.*, *Immunology* 91(1): 13-9 (1997); and Nesbit *et al.*, *J. Immunol. Methods* 151(1-2): 201-8 (1992).

Antibodies and fragments and derivatives thereof of the present invention can also
10 be produced in plant cells, particularly maize or tobacco, Giddings *et al.*, *Nature Biotechnol.* 18(11): 1151-5 (2000); Gavilondo *et al.*, *Biotechniques* 29(1): 128-38 (2000); Fischer *et al.*, *J. Biol. Regul. Homeost. Agents* 14(2): 83-92 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 113-6 (1999); Fischer *et al.*, *Biol. Chem.* 380(7-8): 825-39 (1999); Russell, *Curr. Top. Microbiol. Immunol.* 240: 119-38 (1999); and Ma *et al.*, *Plant Physiol.* 109(2): 341-6 (1995).

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in transgenic, non-human, mammalian milk. See, e.g. Pollock *et al.*, *J. Immunol Methods.* 231: 147-57 (1999); Young *et al.*, *Res. Immunol.* 149: 609-10 (1998); and Limonta *et al.*, *Immunotechnology* 1: 107-13 (1995).

20 Mammalian cells useful for recombinant expression of antibodies, antibody fragments, and antibody derivatives of the present invention include CHO cells, COS cells, 293 cells, and myeloma cells. Verma *et al.*, *J. Immunol. Methods* 216(1-2):165-81 (1998) review and compare bacterial, yeast, insect and mammalian expression systems for expression of antibodies. Antibodies of the present invention can also be prepared by cell
25 free translation, as further described in Merk *et al.*, *J. Biochem. (Tokyo)* 125(2): 328-33 (1999) and Ryabova *et al.*, *Nature Biotechnol.* 15(1): 79-84 (1997), and in the milk of transgenic animals, as further described in Pollock *et al.*, *J. Immunol. Methods* 231(1-2): 147-57 (1999).

The invention further provides antibody fragments that bind specifically to one or
30 more of the polypeptides of the present invention, to one or more of the polypeptides encoded by the isolated nucleic acid molecule of the present invention, or the binding of which can be competitively inhibited by one or more of the polypeptides of the present invention or one or more of the polypeptides encoded by the isolated nucleic acid

molecules of the present invention. Among such useful fragments are Fab, Fab', Fv, F(ab)'₂, and single chain Fv (scFv) fragments. Other useful fragments are described in Hudson, *Curr. Opin. Biotechnol.* 9(4): 395-402 (1998).

5 The present invention also relates to antibody derivatives that bind specifically to one or more of the polypeptides of the present invention, to one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention, or the binding of which can be competitively inhibited by one or more of the polypeptides of the present invention or one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention.

10 Among such useful derivatives are chimeric, primatized, and humanized antibodies; such derivatives are less immunogenic in human beings, and thus are more suitable for *in vivo* administration, than are unmodified antibodies from non-human mammalian species. Another useful derivative is PEGylation to increase the serum half life of the antibodies.

15 Chimeric antibodies typically include heavy and/or light chain variable regions (including both CDR and framework residues) of immunoglobulins of one species, typically mouse, fused to constant regions of another species, typically human. *See, e.g., Morrison et al., Proc. Natl. Acad. Sci USA* 81(21): 6851-5 (1984); Sharon *et al., Nature* 309(5966): 364-7 (1984); Takeda *et al., Nature* 314(6010): 452-4 (1985); and United
20 States Patent No. 5,807,715 the disclosure of which is incorporated herein by reference in its entirety. Primatized and humanized antibodies typically include heavy and/or light chain CDRs from a murine antibody grafted into a non-human primate or human antibody V region framework, usually further comprising a human constant region, Riechmann *et al., Nature* 332(6162): 323-7 (1988); Co *et al., Nature* 351(6326): 501-2 (1991); and
25 United States Patent Nos. 6,054,297; 5,821,337; 5,770,196; 5,766,886; 5,821,123; 5,869,619; 6,180,377; 6,013,256; 5,693,761; and 6,180,370, the disclosures of which are incorporated herein by reference in their entireties. Other useful antibody derivatives of the invention include heteromeric antibody complexes and antibody fusions, such as diabodies (bispecific antibodies), single-chain diabodies, and intrabodies.

30 It is contemplated that the nucleic acids encoding the antibodies of the present invention can be operably joined to other nucleic acids forming a recombinant vector for cloning or for expression of the antibodies of the invention. Accordingly, the present invention includes any recombinant vector containing the coding sequences, or part

thereof, whether for eukaryotic transduction, transfection or gene therapy. Such vectors may be prepared using conventional molecular biology techniques, known to those with skill in the art, and would comprise DNA encoding sequences for the immunoglobulin V-regions including framework and CDRs or parts thereof, and a suitable promoter either
5 with or without a signal sequence for intracellular transport. Such vectors may be transduced or transfected into eukaryotic cells or used for gene therapy (Marasco et al., *Proc. Natl. Acad. Sci. (USA)* 90: 7889-7893 (1993); Duan et al., *Proc. Natl. Acad. Sci. (USA)* 91: 5075-5079 (1994), by conventional techniques, known to those with skill in the art.

10 The antibodies of the present invention, including fragments and derivatives thereof, can usefully be labeled. It is, therefore, another aspect of the present invention to provide labeled antibodies that bind specifically to one or more of the polypeptides of the present invention, to one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention, or the binding of which can be competitively
15 inhibited by one or more of the polypeptides of the present invention or one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention. The choice of label depends, in part, upon the desired use.

For example, when the antibodies of the present invention are used for immunohistochemical staining of tissue samples, the label can usefully be an enzyme that
20 catalyzes production and local deposition of a detectable product. Enzymes typically conjugated to antibodies to permit their immunohistochemical visualization are well-known, and include alkaline phosphatase, β -galactosidase, glucose oxidase, horseradish peroxidase (HRP), and urease. Typical substrates for production and deposition of visually detectable products include o-nitrophenyl-beta-D-galactopyranoside (ONPG);
25 o-phenylenediamine dihydrochloride (OPD); p-nitrophenyl phosphate (PNPP); p-nitrophenyl-beta-D-galactopyranoside (PNPG); 3',3'-diaminobenzidine (DAB); 3-amino-9-ethylcarbazole (AEC); 4-chloro-1-naphthol (CN);
5-bromo-4-chloro-3-indolyl-phosphate (BCIP); ABTS®; BlueGal; iodonitrotetrazolium (INT); nitroblue tetrazolium chloride (NBT); phenazine methosulfate (PMS);
30 phenolphthalein monophosphate (PMP); tetramethyl benzidine (TMB); tetranitroblue tetrazolium (TNBT); X-Gal; X-Gluc; and X-Glucoside.

Other substrates can be used to produce products for local deposition that are luminescent. For example, in the presence of hydrogen peroxide (H₂O₂), horseradish

peroxidase (HRP) can catalyze the oxidation of cyclic diacylhydrazides, such as luminol. Immediately following the oxidation, the luminol is in an excited state (intermediate reaction product), which decays to the ground state by emitting light. Strong enhancement of the light emission is produced by enhancers, such as phenolic compounds. Advantages
5 include high sensitivity, high resolution, and rapid detection without radioactivity and requiring only small amounts of antibody. See, e.g., Thorpe *et al.*, *Methods Enzymol.* 133: 331-53 (1986); Kricka *et al.*, *J. Immunoassay* 17(1): 67-83 (1996); and Lundqvist *et al.*, *J. Biolumin. Chemilumin.* 10(6): 353-9 (1995). Kits for such enhanced chemiluminescent detection (ECL) are available commercially. The antibodies can also be labeled using
10 colloidal gold.

As another example, when the antibodies of the present invention are used, e.g., for flow cytometric detection, for scanning laser cytometric detection, or for fluorescent immunoassay, they can usefully be labeled with fluorophores. There are a wide variety of fluorophore labels that can usefully be attached to the antibodies of the present invention.
15 For flow cytometric applications, both for extracellular detection and for intracellular detection, common useful fluorophores can be fluorescein isothiocyanate (FITC), allophycocyanin (APC), R-phycoerythrin (PE), peridinin chlorophyll protein (PerCP), Texas Red, Cy3, Cy5, fluorescence resonance energy tandem fluorophores such as PerCP-Cy5.5, PE-Cy5, PE-Cy5.5, PE-Cy7, PE-Texas Red, and APC-Cy7.

Other fluorophores include, *inter alia*, Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY
25 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA), and Cy2, Cy3, Cy3.5, Cy5, Cy5.5, Cy7, all of
30 which are also useful for fluorescently labeling the antibodies of the present invention. For secondary detection using labeled avidin, streptavidin, captavidin or neutravidin, the antibodies of the present invention can usefully be labeled with biotin.

When the antibodies of the present invention are used, *e.g.*, for western blotting applications, they can usefully be labeled with radioisotopes, such as ^{33}P , ^{32}P , ^{35}S , ^3H , and ^{125}I . As another example, when the antibodies of the present invention are used for radioimmunotherapy, the label can usefully be ^{228}Th , ^{227}Ac , ^{225}Ac , ^{223}Ra , ^{213}Bi , ^{212}Pb , ^{212}Bi , ^{211}At , ^{203}Pb , ^{194}Os , ^{188}Re , ^{186}Re , ^{153}Sm , ^{149}Tb , ^{131}I , ^{125}I , ^{111}In , ^{105}Rh , $^{99\text{m}}\text{Tc}$, ^{97}Ru , ^{90}Y , ^{90}Sr , ^{88}Y , ^{72}Se , ^{67}Cu , or ^{47}Sc .

As another example, when the antibodies of the present invention are to be used for *in vivo* diagnostic use, they can be rendered detectable by conjugation to MRI contrast agents, such as gadolinium diethylenetriaminepentaacetic acid (DTPA), Lauffer *et al.*, *Radiology* 207(2): 529-38 (1998), or by radioisotopic labeling.

As would be understood, use of the labels described above is not restricted to the application as for which they were mentioned.

The antibodies of the present invention, including fragments and derivatives thereof, can also be conjugated to toxins, in order to target the toxin's ablative action to cells that display and/or express the polypeptides of the present invention. Commonly, the antibody in such immunotoxins is conjugated to Pseudomonas exotoxin A, diphtheria toxin, shiga toxin A, anthrax toxin lethal factor, or ricin. *See* Hall (ed.), Immunotoxin Methods and Protocols (Methods in Molecular Biology, vol. 166), Humana Press (2000); and Frankel *et al.* (eds.), Clinical Applications of Immunotoxins, Springer-Verlag (1998).

The antibodies of the present invention can usefully be attached to a substrate, and it is, therefore, another aspect of the invention to provide antibodies that bind specifically to one or more of the polypeptides of the present invention, to one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention, or the binding of which can be competitively inhibited by one or more of the polypeptides of the present invention or one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention, attached to a substrate. Substrates can be porous or nonporous, planar or nonplanar. For example, the antibodies of the present invention can usefully be conjugated to filtration media, such as NHS-activated Sepharose or CNBr-activated Sepharose for purposes of immunoaffinity chromatography. For example, the antibodies of the present invention can usefully be attached to paramagnetic microspheres, typically by biotin-streptavidin interaction, which microsphere can then be used for isolation of cells that express or display the polypeptides of the present invention. As

another example, the antibodies of the present invention can usefully be attached to the surface of a microtiter plate for ELISA.

As noted above, the antibodies of the present invention can be produced in prokaryotic and eukaryotic cells. It is, therefore, another aspect of the present invention to
5 provide cells that express the antibodies of the present invention, including hybridoma cells, B cells, plasma cells, and host cells recombinantly modified to express the antibodies of the present invention.

In yet a further aspect, the present invention provides aptamers evolved to bind specifically to one or more of the polypeptides of the present invention, to one or more of
10 the polypeptides encoded by the isolated nucleic acid molecules of the present invention, or the binding of which can be competitively inhibited by one or more of the polypeptides of the present invention or one or more of the polypeptides encoded by the isolated nucleic acid molecules of the present invention.

In sum, one of skill in the art, provided with the teachings of this invention, has
15 available a variety of methods which may be used to alter the biological properties of the antibodies of this invention including methods which would increase or decrease the stability or half-life, immunogenicity, toxicity, affinity or yield of a given antibody molecule, or to alter it in any other way that may render it more suitable for a particular application.

20 Transgenic Animals and Cells

In another aspect, the invention provides transgenic cells and non-human organisms comprising nucleic acid molecules of the invention. In a preferred embodiment, the transgenic cells and non-human organisms comprise a nucleic acid
25 molecule encoding a CSP. In a preferred embodiment, the CSP comprises an amino acid sequence selected from SEQ ID NO: 101-190, or a fragment, mutein, homologous protein or allelic variant thereof. In another preferred embodiment, the transgenic cells and non-human organism comprise a CSNA of the invention, preferably a CSNA comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-100, or a part,
30 substantially similar nucleic acid molecule, allelic variant or hybridizing nucleic acid molecule thereof.

In another embodiment, the transgenic cells and non-human organisms have a targeted disruption or replacement of the endogenous orthologue of the human CSG. The

transgenic cells can be embryonic stem cells or somatic cells. The transgenic non-human organisms can be chimeric, nonchimeric heterozygotes, and nonchimeric homozygotes. Methods of producing transgenic animals are well-known in the art. *See, e.g.,* Hogan *et al.*, Manipulating the Mouse Embryo: A Laboratory Manual, 2d ed., Cold Spring Harbor Press (1999); Jackson *et al.*, Mouse Genetics and Transgenics: A Practical Approach, Oxford University Press (2000); and Pinkert, Transgenic Animal Technology: A Laboratory Handbook, Academic Press (1999).

Any technique known in the art may be used to introduce a nucleic acid molecule of the invention into an animal to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection. (*see, e.g.,* Paterson *et al.*, *Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver *et al.*, *Biotechnology* 11: 1263-1270 (1993); Wright *et al.*, *Biotechnology* 9: 830-834 (1991); and United States Patent No. 4,873,191, herein incorporated by reference in its entirety); retrovirus-mediated gene transfer into germ lines, blastocysts or embryos (*see, e.g.,* Van der Putten *et al.*, *Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)); gene targeting in embryonic stem cells (*see, e.g.,* Thompson *et al.*, *Cell* 56: 313-321 (1989)); electroporation of cells or embryos (*see, e.g.,* Lo, 1983, *Mol. Cell. Biol.* 3: 1803-1814 (1983)); introduction using a gene gun (*see, e.g.,* Ulmer *et al.*, *Science* 259: 1745-49 (1993); introducing nucleic acid constructs into embryonic pluripotent stem cells and transferring the stem cells back into the blastocyst; and sperm-mediated gene transfer (*see, e.g.,* Lavitrano *et al.*, *Cell* 57: 717-723 (1989)).

Other techniques include, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (*see, e.g.,* Campell *et al.*, *Nature* 380: 64-66 (1996); Wilmut *et al.*, *Nature* 385: 810-813 (1997)). The present invention provides for transgenic animals that carry the transgene (*i.e.,* a nucleic acid molecule of the invention) in all their cells, as well as animals which carry the transgene in some, but not all their cells, *i.e.,* mosaic animals or chimeric animals.

The transgene may be integrated as a single transgene or as multiple copies, such as in concatamers, *e.g.,* head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, *e.g.,* the teaching of Lasko *et al. et al.*, *Proc. Natl. Acad. Sci. USA* 89: 6232- 6236 (1992). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (RT-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Methods for creating a transgenic animal with a disruption of a targeted gene are also well-known in the art. In general, a vector is designed to comprise some nucleotide sequences homologous to the endogenous targeted gene. The vector is introduced into a cell so that it may integrate, via homologous recombination with chromosomal sequences, into the endogenous gene, thereby disrupting the function of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type. *See, e.g., Gu et al., Science* 265: 103-106

(1994). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. *See, e.g.,* Smithies *et al.*, *Nature* 317: 230-234 (1985); Thomas *et al.*, *Cell* 51: 503-512 (1987); Thompson *et al.*, *Cell* 5: 313-321 (1989).

5 In one embodiment, a mutant, non-functional nucleic acid molecule of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous nucleic acid sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention in vivo. In another
10 embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted
15 gene. *See, e.g.,* Thomas, *supra* and Thompson, *supra*. However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site in vivo using appropriate viral vectors that will be apparent to those of skill in the art.

 In further embodiments of the invention, cells that are genetically engineered to
20 express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (*e.g.*, knockouts) are administered to a patient in vivo. Such cells may be obtained from an animal or patient or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (*e.g.*, lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells
25 are genetically engineered in vitro using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, *e.g.*, by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures,
30 including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

 The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve

expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, *e.g.*, in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, *e.g.*, genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. *See, e.g.*, United States Patent Nos. 5,399,349 and 5,460,959, each of which is incorporated by reference herein in its entirety.

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well-known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Computer Readable Means

A further aspect of the invention is a computer readable means for storing the nucleic acid and amino acid sequences of the instant invention. In a preferred embodiment, the invention provides a computer readable means for storing SEQ ID NO: 101-190 and SEQ ID NO: 1-100 as described herein, as the complete set of sequences or in any combination. The records of the computer readable means can be accessed for reading and display and for interface with a computer system for the application of programs allowing for the location of data upon a query for data meeting certain criteria, the comparison of sequences, the alignment or ordering of sequences meeting a set of criteria, and the like.

The nucleic acid and amino acid sequences of the invention are particularly useful as components in databases useful for search analyses as well as in sequence analysis algorithms. As used herein, the terms "nucleic acid sequences of the invention" and

“amino acid sequences of the invention” mean any detectable chemical or physical characteristic of a polynucleotide or polypeptide of the invention that is or may be reduced to or stored in a computer readable form. These include, without limitation, chromatographic scan data or peak data, photographic data or scan data therefrom, and
5 mass spectrographic data.

This invention provides computer readable media having stored thereon sequences of the invention. A computer readable medium may comprise one or more of the following: a nucleic acid sequence comprising a sequence of a nucleic acid sequence of the invention; an amino acid sequence comprising an amino acid sequence of the
10 invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of one or more nucleic acid sequences of the invention; a data set representing a nucleic
15 acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence
20 comprising the sequence of a nucleic acid sequence of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention. The computer readable medium can be any composition of matter used to store information or data, including, for example, commercially available floppy disks, tapes, hard drives, compact disks, and video disks.

Also provided by the invention are methods for the analysis of character sequences, particularly genetic sequences. Preferred methods of sequence analysis include, for example, methods of sequence homology analysis, such as identity and similarity analysis, RNA structure analysis, sequence assembly, cladistic analysis, sequence motif analysis, open reading frame determination, nucleic acid base calling, and
25 sequencing chromatogram peak analysis.

A computer-based method is provided for performing nucleic acid sequence identity or similarity identification. This method comprises the steps of providing a nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a

computer readable medium; and comparing said nucleic acid sequence to at least one nucleic acid or amino acid sequence to identify sequence identity or similarity.

A computer-based method is also provided for performing amino acid homology identification, said method comprising the steps of: providing an amino acid sequence
5 comprising the sequence of an amino acid of the invention in a computer readable medium; and comparing said amino acid sequence to at least one nucleic acid or an amino acid sequence to identify homology.

A computer-based method is still further provided for assembly of overlapping nucleic acid sequences into a single nucleic acid sequence, said method comprising the
10 steps of: providing a first nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and screening for at least one overlapping region between said first nucleic acid sequence and a second nucleic acid sequence. In addition, the invention includes a method of using patterns of expression associated with either the nucleic acids or proteins in a computer-based method to
15 diagnose disease.

Diagnostic Methods for Colon Cancer

The present invention also relates to quantitative and qualitative diagnostic assays and methods for detecting, diagnosing, monitoring, staging and predicting cancers by
20 comparing expression of a CSNA or a CSP in a human patient that has or may have colon cancer, or who is at risk of developing colon cancer, with the expression of a CSNA or a CSP in a normal human control. For purposes of the present invention, "expression of a CSNA" or "CSNA expression" means the quantity of CSG mRNA that can be measured by any method known in the art or the level of transcription that can be measured by any
25 method known in the art in a cell, tissue, organ or whole patient. Similarly, the term "expression of a CSP" or "CSP expression" means the amount of CSP that can be measured by any method known in the art or the level of translation of a CSNA that can be measured by any method known in the art.

The present invention provides methods for diagnosing colon cancer in a patient,
30 in particular squamous cell carcinoma, by analyzing for changes in levels of CSNA or CSP in cells, tissues, organs or bodily fluids compared with levels of CSNA or CSP in cells, tissues, organs or bodily fluids of preferably the same type from a normal human control, wherein an increase, or decrease in certain cases, in levels of a CSNA or CSP in

the patient versus the normal human control is associated with the presence of colon cancer or with a predilection to the disease. In another preferred embodiment, the present invention provides methods for diagnosing colon cancer in a patient by analyzing changes in the structure of the mRNA of an CSG compared to the mRNA from a normal control.

- 5 These changes include, without limitation, aberrant splicing, alterations in polyadenylation and/or alterations in 5' nucleotide capping. In yet another preferred embodiment, the present invention provides methods for diagnosing colon cancer in a patient by analyzing changes in a CSP compared to a CSP from a normal patient. These changes include, *e.g.*, alterations, including post translational modifications such as glycosylation and/or
- 10 phosphorylation of the CSP or changes in the subcellular CSP localization.
- For purposes of the present invention, diagnosing means that CSNA or CSP levels are used to determine the presence or absence of disease in a patient. As will be understood by those of skill in the art, measurement of other diagnostic parameters may be required for definitive diagnosis or determination of the appropriate treatment for the disease. The
- 15 determination may be made by a clinician, a doctor, a testing laboratory, or a patient using an over the counter test. The patient may have symptoms of disease or may be asymptomatic. In addition, the CSNA or CSP levels of the present invention may be used as screening marker to determine whether further tests or biopsies are warranted. In addition, the CSNA or CSP levels may be used to determine the vulnerability or
- 20 susceptibility to disease.

- In a preferred embodiment, the expression of a CSNA is measured by determining the amount of a mRNA that encodes an amino acid sequence selected from SEQ ID NO: 101-190, a homolog, an allelic variant, or a fragment thereof. In a more preferred embodiment, the CSNA expression that is measured is the level of expression of a CSNA
- 25 mRNA selected from SEQ ID NO: 1-100, or a hybridizing nucleic acid, homologous nucleic acid or allelic variant thereof, or a part of any of these nucleic acid molecules. CSNA expression may be measured by any method known in the art, such as those described *supra*, including measuring mRNA expression by Northern blot, quantitative or qualitative reverse transcriptase PCR (RT-PCR), microarray, dot or slot blots or *in situ*
- 30 hybridization. *See, e.g.*, Ausubel (1992), *supra*; Ausubel (1999), *supra*; Sambrook (1989), *supra*; and Sambrook (2001), *supra*. CSNA transcription may be measured by any method known in the art including using a reporter gene hooked up to the promoter of a CSG of interest or doing nuclear run-off assays. Alterations in mRNA structure, *e.g.*,

aberrant splicing variants, may be determined by any method known in the art, including, RT-PCR followed by sequencing or restriction analysis. As necessary, CSNA expression may be compared to a known control, such as normal colon nucleic acid, to detect a change in expression.

5 In another preferred embodiment, the expression of a CSP is measured by determining the level of a CSP having an amino acid sequence selected from the group consisting of SEQ ID NO: 101-190, a homolog, an allelic variant, or a fragment thereof. Such levels are preferably determined in at least one of cells, tissues, organs and/or bodily fluids, including determination of normal and abnormal levels. Thus, for instance, a
10 diagnostic assay in accordance with the invention for diagnosing over- or underexpression of a CSNA or CSP compared to normal control bodily fluids, cells, or tissue samples may be used to diagnose the presence of colon cancer. The expression level of a CSP may be determined by any method known in the art, such as those described *supra*. In a preferred embodiment, the CSP expression level may be determined by radioimmunoassays,
15 competitive-binding assays, ELISA, Western blot, FACS, immunohistochemistry, immunoprecipitation, proteomic approaches: two-dimensional gel electrophoresis (2D electrophoresis) and non-gel-based approaches such as mass spectrometry or protein interaction profiling. *See, e.g.,* Harlow (1999), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. Alterations in the CSP structure may be determined by any method known
20 in the art, including, *e.g.,* using antibodies that specifically recognize phosphoserine, phosphothreonine or phosphotyrosine residues, two-dimensional polyacrylamide gel electrophoresis (2D PAGE) and/or chemical analysis of amino acid residues of the protein.
Id.

 In a preferred embodiment, a radioimmunoassay (RIA) or an ELISA is used. An
25 antibody specific to a CSP is prepared if one is not already available. In a preferred embodiment, the antibody is a monoclonal antibody. The anti-CSP antibody is bound to a solid support and any free protein binding sites on the solid support are blocked with a protein such as bovine serum albumin. A sample of interest is incubated with the antibody on the solid support under conditions in which the CSP will bind to the anti-CSP antibody.
30 The sample is removed, the solid support is washed to remove unbound material, and an anti-CSP antibody that is linked to a detectable reagent (a radioactive substance for RIA and an enzyme for ELISA) is added to the solid support and incubated under conditions in which binding of the CSP to the labeled antibody will occur. After binding, the unbound

labeled antibody is removed by washing. For an ELISA, one or more substrates are added to produce a colored reaction product that is based upon the amount of an CSP in the sample. For an RIA, the solid support is counted for radioactive decay signals by any method known in the art. Quantitative results for both RIA and ELISA typically are
5 obtained by reference to a standard curve.

Other methods to measure CSP levels are known in the art. For instance, a competition assay may be employed wherein an anti-CSP antibody is attached to a solid support and an allocated amount of a labeled CSP and a sample of interest are incubated with the solid support. The amount of labeled CSP detected which is attached to the solid
10 support can be correlated to the quantity of a CSP in the sample.

Of the proteomic approaches, 2D PAGE is a well-known technique. Isolation of individual proteins from a sample such as serum is accomplished using sequential separation of proteins by isoelectric point and molecular weight. Typically, polypeptides are first separated by isoelectric point (the first dimension) and then separated by size
15 using an electric current (the second dimension). In general, the second dimension is perpendicular to the first dimension. Because no two proteins with different sequences are identical on the basis of both size and charge, the result of 2D PAGE is a roughly square gel in which each protein occupies a unique spot. Analysis of the spots with chemical or antibody probes, or subsequent protein microsequencing can reveal the relative abundance
20 of a given protein and the identity of the proteins in the sample.

Expression levels of a CSNA can be determined by any method known in the art, including PCR and other nucleic acid methods, such as ligase chain reaction (LCR) and nucleic acid sequence based amplification (NASBA), can be used to detect malignant cells for diagnosis and monitoring of various malignancies. For example, reverse-transcriptase
25 PCR (RT-PCR) is a powerful technique which can be used to detect the presence of a specific mRNA population in a complex mixture of thousands of other mRNA species. In RT-PCR, an mRNA species is first reverse transcribed to complementary DNA (cDNA) with use of the enzyme reverse transcriptase; the cDNA is then amplified as in a standard PCR reaction.

30 Hybridization to specific DNA molecules (*e.g.*, oligonucleotides) arrayed on a solid support can be used to both detect the expression of and quantitate the level of expression of one or more CSNAs of interest. In this approach, all or a portion of one or more CSNAs is fixed to a substrate. A sample of interest, which may comprise RNA, *e.g.*,

total RNA or polyA-selected mRNA, or a complementary DNA (cDNA) copy of the RNA is incubated with the solid support under conditions in which hybridization will occur between the DNA on the solid support and the nucleic acid molecules in the sample of interest. Hybridization between the substrate-bound DNA and the nucleic acid molecules
5 in the sample can be detected and quantitated by several means, including, without limitation, radioactive labeling or fluorescent labeling of the nucleic acid molecule or a secondary molecule designed to detect the hybrid.

The above tests can be carried out on samples derived from a variety of cells, bodily fluids and/or tissue extracts such as homogenates or solubilized tissue obtained
10 from a patient. Tissue extracts are obtained routinely from tissue biopsy and autopsy material. Bodily fluids useful in the present invention include blood, urine, saliva or any other bodily secretion or derivative thereof. As used herein "blood" includes whole blood, plasma, serum or any derivative of blood. In a preferred embodiment, the specimen tested for expression of CSNA or CSP includes, without limitation, colon tissue, fluid
15 obtained by bronchial alveolar lavage (BAL), sputum, colon cells grown in cell culture, blood, serum, lymph node tissue and lymphatic fluid. In another preferred embodiment, especially when metastasis of a primary colon cancer is known or suspected, specimens include, without limitation, tissues from brain, bone, bone marrow, liver, adrenal glands and colon. In general, the tissues may be sampled by biopsy, including, without
20 limitation, needle biopsy, *e.g.*, transthoracic needle aspiration, cervical mediastinoscopy, endoscopic lymph node biopsy, video-assisted thoracoscopy, exploratory thoracotomy, bone marrow biopsy and bone marrow aspiration. *See Scott, supra* and Franklin, pp. 529-570, in Kane, *supra*. For early and inexpensive detection, assaying for changes in CSNAs or CSPs in cells in sputum samples may be particularly useful. Methods of obtaining and
25 analyzing sputum samples are disclosed in Franklin, *supra*.

All the methods of the present invention may optionally include determining the expression levels of one or more other cancer markers in addition to determining the expression level of a CSNA or CSP. In many cases, the use of another cancer marker will decrease the likelihood of false positives or false negatives. In one embodiment, the one or
30 more other cancer markers include other CSNA or CSPs as disclosed herein. Other cancer markers useful in the present invention will depend on the cancer being tested and are known to those of skill in the art. In a preferred embodiment, at least one other cancer marker in addition to a particular CSNA or CSP is measured. In a more preferred

embodiment, at least two other additional cancer markers are used. In an even more preferred embodiment, at least three, more preferably at least five, even more preferably at least ten additional cancer markers are used.

Diagnosing

5 In one aspect, the invention provides a method for determining the expression levels and/or structural alterations of one or more CSNA and/or CSP in a sample from a patient suspected of having colon cancer. In general, the method comprises the steps of obtaining the sample from the patient, determining the expression level or structural alterations of a CSNA and/or CSP and then ascertaining whether the patient has colon
10 cancer from the expression level of the CSNA or CSP. In general, if high expression relative to a control of a CSNA or CSP is indicative of colon cancer, a diagnostic assay is considered positive if the level of expression of the CSNA or CSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human
15 control. In contrast, if low expression relative to a control of a CSNA or CSP is indicative of colon cancer, a diagnostic assay is considered positive if the level of expression of the CSNA or CSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a
20 different patient or from uninvolved tissue of the same patient.

The present invention also provides a method of determining whether colon cancer has metastasized in a patient. One may identify whether the colon cancer has metastasized by measuring the expression levels and/or structural alterations of one or more CSNAs and/or CSPs in a variety of tissues. The presence of a CSNA or CSP in a certain tissue at
25 levels higher than that of corresponding noncancerous tissue (*e.g.*, the same tissue from another individual) is indicative of metastasis if high level expression of a CSNA or CSP is associated with colon cancer. Similarly, the presence of a CSNA or CSP in a tissue at levels lower than that of corresponding noncancerous tissue is indicative of metastasis if low level expression of a CSNA or CSP is associated with colon cancer. Further, the
30 presence of a structurally altered CSNA or CSP that is associated with colon cancer is also indicative of metastasis.

In general, if high expression relative to a control of n CSNA or CSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of

the CSNA or CSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a CSNA or CSP is indicative of metastasis, an assay for metastasis is

5 considered positive if the level of expression of the CSNA or CSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control.

Staging

10 The invention also provides a method of staging colon cancer in a human patient. The method comprises identifying a human patient having colon cancer and analyzing cells, tissues or bodily fluids from such human patient for expression levels and/or structural alterations of one or more CSNAs or CSPs. First, one or more tumors from a variety of patients are staged according to procedures well-known in the art, and the

15 expression levels of one or more CSNAs or CSPs is determined for each stage to obtain a standard expression level for each CSNA and CSP. Then, the CSNA or CSP expression levels of the CSNA or CSP are determined in a biological sample from a patient whose stage of cancer is not known. The CSNA or CSP expression levels from the patient are then compared to the standard expression level. By comparing the expression level of the

20 CSNAs and CSPs from the patient to the standard expression levels, one may determinethe stage of the tumor. The same procedure may be followed using structural alterations of a CSNA or CSP to determine the stage of a colon cancer.

Monitoring

Further provided is a method of monitoring colon cancer in a human patient. One

25 may monitor a human patient to determine whether there has been metastasis and, if there has been, when metastasis began to occur. One may also monitor a human patient to determine whether a preneoplastic lesion has become cancerous. One may also monitor a human patient to determine whether a therapy, e.g., chemotherapy, radiotherapy or surgery, has decreased or eliminated the colon cancer. The monitoring may determine if

30 there has been a reoccurrence and, if so, determine its nature. The method comprises identifying a human patient that one wants to monitor for colon cancer, periodically analyzing cells, tissues or bodily fluids from such human patient for expression levels of

one or more CSNAs or CSPs, and comparing the CSNA or CSP levels over time to those CSNA or CSP expression levels obtained previously. Patients may also be monitored by measuring one or more structural alterations in a CSNA or CSP that are associated with colon cancer.

5 If increased expression of a CSNA or CSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an increase in the expression level of a CSNA or CSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. One having ordinary skill in the art would recognize that if this were the case, then a decreased
10 expression level would be indicative of no metastasis, effective therapy or failure to progress to a neoplastic lesion. If decreased expression of a CSNA or CSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an decrease in the expression level of a CSNA or CSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous,
15 respectively. In a preferred embodiment, the levels of CSNAs or CSPs are determined from the same cell type, tissue or bodily fluid as prior patient samples. Monitoring a patient for onset of colon cancer metastasis is periodic and preferably is done on a quarterly basis, but may be done more or less frequently.

 The methods described herein can further be utilized as prognostic assays to
20 identify subjects having or at risk of developing a disease or disorder associated with increased or decreased expression levels of a CSNA and/or CSP. The present invention provides a method in which a test sample is obtained from a human patient and one or more CSNAs and/or CSPs are detected. The presence of higher (or lower) CSNA or CSP levels as compared to normal human controls is diagnostic for the human patient being at
25 risk for developing cancer, particularly colon cancer. The effectiveness of therapeutic agents to decrease (or increase) expression or activity of one or more CSNAs and/or CSPs of the invention can also be monitored by analyzing levels of expression of the CSNAs and/or CSPs in a human patient in clinical trials or in *in vitro* screening assays such as in human cells. In this way, the gene expression pattern can serve as a marker, indicative of
30 the physiological response of the human patient or cells, as the case may be, to the agent being tested.

Detection of Genetic Lesions or Mutations

The methods of the present invention can also be used to detect genetic lesions or mutations in a CSG, thereby determining if a human with the genetic lesion is susceptible to developing colon cancer or to determine what genetic lesions are responsible, or are partly responsible, for a person's existing colon cancer. Genetic lesions can be detected, for example, by ascertaining the existence of a deletion, insertion and/or substitution of one or more nucleotides from the CSGs of this invention, a chromosomal rearrangement of a CSG, an aberrant modification of a CSG (such as of the methylation pattern of the genomic DNA), or allelic loss of a CSG. Methods to detect such lesions in the CSG of this invention are known to those having ordinary skill in the art following the teachings of the specification.

Methods of Detecting Noncancerous Colon Diseases

The present invention also provides methods for determining the expression levels and/or structural alterations of one or more CSNAs and/or CSPs in a sample from a patient suspected of having or known to have a noncancerous colon disease. In general, the method comprises the steps of obtaining a sample from the patient, determining the expression level or structural alterations of a CSNA and/or CSP, comparing the expression level or structural alteration of the CSNA or CSP to a normal colon control, and then ascertaining whether the patient has a noncancerous colon disease. In general, if high expression relative to a control of a CSNA or CSP is indicative of a particular noncancerous colon disease, a diagnostic assay is considered positive if the level of expression of the CSNA or CSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a CSNA or CSP is indicative of a noncancerous colon disease, a diagnostic assay is considered positive if the level of expression of the CSNA or CSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

One having ordinary skill in the art may determine whether a CSNA and/or CSP is associated with a particular noncancerous colon disease by obtaining colon tissue from a patient having a noncancerous colon disease of interest and determining which CSNAs

and/or CSPs are expressed in the tissue at either a higher or a lower level than in normal colon tissue. In another embodiment, one may determine whether a CSNA or CSP exhibits structural alterations in a particular noncancerous colon disease state by obtaining colon tissue from a patient having a noncancerous colon disease of interest and
5 determining the structural alterations in one or more CSNAs and/or CSPs relative to normal colon tissue.

Methods for Identifying Colon Tissue

In another aspect, the invention provides methods for identifying colon tissue.
10 These methods are particularly useful in, *e.g.*, forensic science, colon cell differentiation and development, and in tissue engineering.

In one embodiment, the invention provides a method for determining whether a sample is colon tissue or has colon tissue-like characteristics. The method comprises the steps of providing a sample suspected of comprising colon tissue or having colon tissue-
15 like characteristics, determining whether the sample expresses one or more CSNAs and/or CSPs, and, if the sample expresses one or more CSNAs and/or CSPs, concluding that the sample comprises colon tissue. In a preferred embodiment, the CSNA encodes a polypeptide having an amino acid sequence selected from SEQ ID NO: 101-190, or a homolog, allelic variant or fragment thereof. In a more preferred embodiment, the CSNA
20 has a nucleotide sequence selected from SEQ ID NO: 1-100, or a hybridizing nucleic acid, an allelic variant or a part thereof. Determining whether a sample expresses a CSNA can be accomplished by any method known in the art. Preferred methods include hybridization to microarrays, Northern blot hybridization, and quantitative or qualitative RT-PCR. In another preferred embodiment, the method can be practiced by determining
25 whether a CSP is expressed. Determining whether a sample expresses a CSP can be accomplished by any method known in the art. Preferred methods include Western blot, ELISA, RIA and 2D PAGE. In one embodiment, the CSP has an amino acid sequence selected from SEQ ID NO: 101-190, or a homolog, allelic variant or fragment thereof. In another preferred embodiment, the expression of at least two CSNAs and/or CSPs is
30 determined. In a more preferred embodiment, the expression of at least three, more preferably four and even more preferably five CSNAs and/or CSPs are determined.

In one embodiment, the method can be used to determine whether an unknown tissue is colon tissue. This is particularly useful in forensic science, in which small,

damaged pieces of tissues that are not identifiable by microscopic or other means are recovered from a crime or accident scene. In another embodiment, the method can be used to determine whether a tissue is differentiating or developing into colon tissue. This is important in monitoring the effects of the addition of various agents to cell or tissue culture, *e.g.*, in producing new colon tissue by tissue engineering. These agents include, *e.g.*, growth and differentiation factors, extracellular matrix proteins and culture medium. Other factors that may be measured for effects on tissue development and differentiation include gene transfer into the cells or tissues, alterations in pH, aqueous:air interface and various other culture conditions.

10 Methods for Producing and Modifying Colon Tissue

In another aspect, the invention provides methods for producing engineered colon tissue or cells. In one embodiment, the method comprises the steps of providing cells, introducing a CSNA or a CSG into the cells, and growing the cells under conditions in which they exhibit one or more properties of colon tissue cells. In a preferred embodiment, the cells are pluripotent. As is well-known in the art, normal colon tissue comprises a large number of different cell types. Thus, in one embodiment, the engineered colon tissue or cells comprises one of these cell types. In another embodiment, the engineered colon tissue or cells comprises more than one colon cell type. Further, the culture conditions of the cells or tissue may require manipulation in order to achieve full differentiation and development of the colon cell tissue. Methods for manipulating culture conditions are well-known in the art.

Nucleic acid molecules encoding one or more CSPs are introduced into cells, preferably pluripotent cells. In a preferred embodiment, the nucleic acid molecules encode CSPs having amino acid sequences selected from SEQ ID NO: 101-190, or homologous proteins, analogs, allelic variants or fragments thereof. In a more preferred embodiment, the nucleic acid molecules have a nucleotide sequence selected from SEQ ID NO: 1-100, or hybridizing nucleic acids, allelic variants or parts thereof. In another highly preferred embodiment, a CSG is introduced into the cells. Expression vectors and methods of introducing nucleic acid molecules into cells are well-known in the art and are described in detail, *supra*.

Artificial colon tissue may be used to treat patients who have lost some or all of their colon function.

Pharmaceutical Compositions

5 In another aspect, the invention provides pharmaceutical compositions comprising the nucleic acid molecules, polypeptides, fusion proteins, antibodies, antibody derivatives, antibody fragments, agonists, antagonists, or inhibitors of the present invention. In a preferred embodiment, the pharmaceutical composition comprises a CSNA or part thereof. In a more preferred embodiment, the CSNA has a nucleotide
10 sequence selected from the group consisting of SEQ ID NO: 1-100, a nucleic acid that hybridizes thereto, an allelic variant thereof, or a nucleic acid that has substantial sequence identity thereto. In another preferred embodiment, the pharmaceutical composition comprises a CSP or fragment thereof. In a more preferred embodiment, the pharmaceutical composition comprises a CSP having an amino acid sequence that is
15 selected from the group consisting of SEQ ID NO: 101-190, a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof. In another preferred embodiment, the pharmaceutical composition comprises an anti-CSP antibody, preferably an antibody that specifically binds to a CSP having an amino acid that is selected from the group consisting of SEQ ID
20 NO: 101-190, or an antibody that binds to a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof.

Such a composition typically contains from about 0.1 to 90% by weight of a therapeutic agent of the invention formulated in and/or with a pharmaceutically acceptable
25 carrier or excipient.

Pharmaceutical formulation is a well-established art that is further described in Gennaro (ed.), Remington: The Science and Practice of Pharmacy, 20th ed., Lippincott, Williams & Wilkins (2000); Ansel *et al.*, Pharmaceutical Dosage Forms and Drug Delivery Systems, 7th ed., Lippincott Williams & Wilkins (1999); and Kibbe (ed.),
30 Handbook of Pharmaceutical Excipients American Pharmaceutical Association, 3rd ed. (2000) and thus need not be described in detail herein.

Briefly, formulation of the pharmaceutical compositions of the present invention will depend upon the route chosen for administration. The pharmaceutical compositions

utilized in this invention can be administered by various routes including both enteral and parenteral routes, including oral, intravenous, intramuscular, subcutaneous, inhalation, topical, sublingual, rectal, intra-arterial, intramedullary, intrathecal, intraventricular, transmucosal, transdermal, intranasal, intraperitoneal, intrapulmonary, and intrauterine.

5 Oral dosage forms can be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

 Solid formulations of the compositions for oral administration can contain suitable carriers or excipients, such as carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other
10 plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, or microcrystalline cellulose; gums including arabic and tragacanth; proteins such as gelatin and collagen; inorganics, such as kaolin, calcium carbonate, dicalcium phosphate, sodium chloride; and other agents such as acacia and alginic acid.

15 Agents that facilitate disintegration and/or solubilization can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate, microcrystalline cellulose, corn starch, sodium starch glycolate, and alginic acid.

 Tablet binders that can be used include acacia, methylcellulose, sodium carboxymethylcellulose, polyvinylpyrrolidone (Povidone™), hydroxypropyl
20 methylcellulose, sucrose, starch and ethylcellulose.

 Lubricants that can be used include magnesium stearates, stearic acid, silicone fluid, talc, waxes, oils, and colloidal silica.

 Fillers, agents that facilitate disintegration and/or solubilization, tablet binders and lubricants, including the aforementioned, can be used singly or in combination.

25 Solid oral dosage forms need not be uniform throughout. For example, dragee cores can be used in conjunction with suitable coatings, such as concentrated sugar solutions, which can also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures.

30 Oral dosage forms of the present invention include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally,

stabilizers. In soft capsules, the active compounds can be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Additionally, dyestuffs or pigments can be added to the tablets or dragee coatings
5 for product identification or to characterize the quantity of active compound, *i.e.*, dosage.

Liquid formulations of the pharmaceutical compositions for oral (enteral) administration are prepared in water or other aqueous vehicles and can contain various suspending agents such as methylcellulose, alginates, tragacanth, pectin, kelgin, carrageenan, acacia, polyvinylpyrrolidone, and polyvinyl alcohol. The liquid formulations
10 can also include solutions, emulsions, syrups and elixirs containing, together with the active compound(s), wetting agents, sweeteners, and coloring and flavoring agents.

The pharmaceutical compositions of the present invention can also be formulated for parenteral administration. Formulations for parenteral administration can be in the form of aqueous or non-aqueous isotonic sterile injection solutions or suspensions.

15 For intravenous injection, water soluble versions of the compounds of the present invention are formulated in, or if provided as a lyophilate, mixed with, a physiologically acceptable fluid vehicle, such as 5% dextrose ("D5"), physiologically buffered saline, 0.9% saline, Hanks' solution, or Ringer's solution. Intravenous formulations may include carriers, excipients or stabilizers including, without limitation, calcium, human serum
20 albumin, citrate, acetate, calcium chloride, carbonate, and other salts.

Intramuscular preparations, *e.g.* a sterile formulation of a suitable soluble salt form of the compounds of the present invention, can be dissolved and administered in a pharmaceutical excipient such as Water-for-Injection, 0.9% saline, or 5% glucose solution. Alternatively, a suitable insoluble form of the compound can be prepared and
25 administered as a suspension in an aqueous base or a pharmaceutically acceptable oil base, such as an ester of a long chain fatty acid (*e.g.*, ethyl oleate), fatty oils such as sesame oil, triglycerides, or liposomes.

Parenteral formulations of the compositions can contain various carriers such as vegetable oils, dimethylacetamide, dimethylformamide, ethyl lactate, ethyl carbonate,
30 isopropyl myristate, ethanol, polyols (glycerol, propylene glycol, liquid polyethylene glycol, and the like).

Aqueous injection suspensions can also contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran.

Non-lipid polycationic amino polymers can also be used for delivery. Optionally, the suspension can also contain suitable stabilizers or agents that increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical compositions of the present invention can also be formulated to
5 permit injectable, long-term, deposition. Injectable depot forms may be made by forming microencapsulated matrices of the compound in biodegradable polymers such as polylactide-polyglycolide. Depending upon the ratio of drug to polymer and the nature of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot
10 injectable formulations are also prepared by entrapping the drug in microemulsions that are compatible with body tissues.

The pharmaceutical compositions of the present invention can be administered topically. For topical use the compounds of the present invention can also be prepared in suitable forms to be applied to the skin, or mucus membranes of the nose and throat, and
15 can take the form of lotions, creams, ointments, liquid sprays or inhalants, drops, tinctures, lozenges, or throat paints. Such topical formulations further can include chemical compounds such as dimethylsulfoxide (DMSO) to facilitate surface penetration of the active ingredient. In other transdermal formulations, typically in patch-delivered formulations, the pharmaceutically active compound is formulated with one or more skin
20 penetrants, such as 2-N-methyl-pyrrolidone (NMP) or Azone. A topical semi-solid ointment formulation typically contains a concentration of the active ingredient from about 1 to 20%, *e.g.*, 5 to 10%, in a carrier such as a pharmaceutical cream base.

For application to the eyes or ears, the compounds of the present invention can be presented in liquid or semi-liquid form formulated in hydrophobic or hydrophilic bases as
25 ointments, creams, lotions, paints or powders.

For rectal administration the compounds of the present invention can be administered in the form of suppositories admixed with conventional carriers such as cocoa butter, wax or other glyceride.

Inhalation formulations can also readily be formulated. For inhalation, various
30 powder and liquid formulations can be prepared. For aerosol preparations, a sterile formulation of the compound or salt form of the compound may be used in inhalers, such as metered dose inhalers, and nebulizers. Aerosolized forms may be especially useful for treating respiratory disorders.

Alternatively, the compounds of the present invention can be in powder form for reconstitution in the appropriate pharmaceutically acceptable carrier at the time of delivery.

5 The pharmaceutically active compound in the pharmaceutical compositions of the present invention can be provided as the salt of a variety of acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms.

10 After pharmaceutical compositions have been prepared, they are packaged in an appropriate container and labeled for treatment of an indicated condition.

The active compound will be present in an amount effective to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

15 A "therapeutically effective dose" refers to that amount of active ingredient, for example CSP polypeptide, fusion protein, or fragments thereof, antibodies specific for CSP, agonists, antagonists or inhibitors of CSP, which ameliorates the signs or symptoms of the disease or prevent progression thereof; as would be understood in the medical arts, cure, although desired, is not required.

20 The therapeutically effective dose of the pharmaceutical agents of the present invention can be estimated initially by *in vitro* tests, such as cell culture assays, followed by assay in model animals, usually mice, rats, rabbits, dogs, or pigs. The animal model can also be used to determine an initial preferred concentration range and route of administration.

25 For example, the ED₅₀ (the dose therapeutically effective in 50% of the population) and LD₅₀ (the dose lethal to 50% of the population) can be determined in one or more cell culture of animal model systems. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as LD₅₀/ED₅₀. Pharmaceutical compositions that exhibit large therapeutic indices are preferred.

30 The data obtained from cell culture assays and animal studies are used in formulating an initial dosage range for human use, and preferably provide a range of circulating concentrations that includes the ED₅₀ with little or no toxicity. After administration, or between successive administrations, the circulating concentration of active agent varies within this range depending upon pharmacokinetic factors well-known

in the art, such as the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors specific to the subject requiring treatment. Factors that can be taken into account by the practitioner include the severity of the disease state, general health of the subject, age, weight, gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Where the therapeutic agent is a protein or antibody of the present invention, the therapeutic protein or antibody agent typically is administered at a daily dosage of 0.01 mg to 30 mg/kg of body weight of the patient (*e.g.*, 1 mg/kg to 5 mg/kg). The pharmaceutical formulation can be administered in multiple doses per day, if desired, to achieve the total desired daily dose.

Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

Conventional methods, known to those of ordinary skill in the art of medicine, can be used to administer the pharmaceutical formulation(s) of the present invention to the patient. The pharmaceutical compositions of the present invention can be administered alone, or in combination with other therapeutic agents or interventions.

Therapeutic Methods

The present invention further provides methods of treating subjects having defects in a gene of the invention, *e.g.*, in expression, activity, distribution, localization, and/or solubility, which can manifest as a disorder of colon function. As used herein, "treating" includes all medically-acceptable types of therapeutic intervention, including palliation and prophylaxis (prevention) of disease. The term "treating" encompasses any improvement of a disease, including minor improvements. These methods are discussed below.

Gene Therapy and Vaccines

The isolated nucleic acids of the present invention can also be used to drive *in vivo* expression of the polypeptides of the present invention. *In vivo* expression can be driven from a vector, typically a viral vector, often a vector based upon a replication incompetent retrovirus, an adenovirus, or an adeno-associated virus (AAV), for the purpose of gene therapy. *In vivo* expression can also be driven from signals endogenous to the nucleic acid or from a vector, often a plasmid vector, such as pVAX1 (Invitrogen, Carlsbad, CA, USA), for purpose of “naked” nucleic acid vaccination, as further described in United States Patent Nos. 5,589,466; 5,679,647; 5,804,566; 5,830,877; 5,843,913; 5,880,104; 5,958,891; 5,985,847; 6,017,897; 6,110,898; 6,204,250, the disclosures of which are incorporated herein by reference in their entireties. For cancer therapy, it is preferred that the vector also be tumor-selective. See, e.g., Doronin *et al.*, *J. Virol.* 75: 3314-24 (2001).

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a nucleic acid molecule of the present invention is administered. The nucleic acid molecule can be delivered in a vector that drives expression of a CSP, fusion protein, or fragment thereof, or without such vector. Nucleic acid compositions that can drive expression of a CSP are administered, for example, to complement a deficiency in the native CSP, or as DNA vaccines. Expression vectors derived from virus, replication deficient retroviruses, adenovirus, adeno-associated (AAV) virus, herpes virus, or vaccinia virus can be used as can plasmids. See, e.g., Cid-Arregui, *supra*. In a preferred embodiment, the nucleic acid molecule encodes a CSP having the amino acid sequence of SEQ ID NO: 101-190, or a fragment, fusion protein, allelic variant or homolog thereof.

In still other therapeutic methods of the present invention, pharmaceutical compositions comprising host cells that express a CSP, fusions, or fragments thereof can be administered. In such cases, the cells are typically autologous, so as to circumvent xenogeneic or allotypic rejection, and are administered to complement defects in CSP production or activity. In a preferred embodiment, the nucleic acid molecules in the cells encode a CSP having the amino acid sequence of SEQ ID NO: 101-190, or a fragment, fusion protein, allelic variant or homolog thereof.

Antisense Administration

Antisense nucleic acid compositions, or vectors that drive expression of a CSG antisense nucleic acid, are administered to downregulate transcription and/or translation of

a CSG in circumstances in which excessive production, or production of aberrant protein, is the pathophysiologic basis of disease.

Antisense compositions useful in therapy can have a sequence that is complementary to coding or to noncoding regions of a CSG. For example,

5 oligonucleotides derived from the transcription initiation site, *e.g.*, between positions -10 and +10 from the start site, are preferred.

Catalytic antisense compositions, such as ribozymes, that are capable of sequence-specific hybridization to CSG transcripts, are also useful in therapy. *See, e.g.*, Phylactou, *Adv. Drug Deliv. Rev.* 44(2-3): 97-108 (2000); Phylactou *et al.*, *Hum. Mol.*

10 *Genet.* 7(10): 1649-53 (1998); Rossi, *Ciba Found. Symp.* 209: 195-204 (1997); and Sigurdsson *et al.*, *Trends Biotechnol.* 13(8): 286-9 (1995).

Other nucleic acids useful in the therapeutic methods of the present invention are those that are capable of triplex helix formation in or near the CSG genomic locus. Such triplexing oligonucleotides are able to inhibit transcription. *See, e.g.*, Intody *et al.*, *Nucleic*

15 *Acids Res.* 28(21): 4283-90 (2000); and McGuffie *et al.*, *Cancer Res.* 60(14): 3790-9 (2000). Pharmaceutical compositions comprising such triplex forming oligos (TFOs) are administered in circumstances in which excessive production, or production of aberrant protein, is a pathophysiologic basis of disease.

In a preferred embodiment, the antisense molecule is derived from a nucleic acid molecule encoding a CSP, preferably a CSP comprising an amino acid sequence of SEQ ID NO: 101-190, or a fragment, allelic variant or homolog thereof. In a more preferred embodiment, the antisense molecule is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-100, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

25 *Polypeptide Administration*

In one embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a CSP, a fusion protein, fragment, analog or derivative thereof is administered to a subject with a clinically-significant CSP defect.

30 Protein compositions are administered, for example, to complement a deficiency in native CSP. In other embodiments, protein compositions are administered as a vaccine to elicit a humoral and/or cellular immune response to CSP. The immune response can be used to modulate activity of CSP or, depending on the immunogen, to immunize against

aberrant or aberrantly expressed forms, such as mutant or inappropriately expressed isoforms. In yet other embodiments, protein fusions having a toxic moiety are administered to ablate cells that aberrantly accumulate CSP.

In a preferred embodiment, the polypeptide administered is a CSP comprising an amino acid sequence of SEQ ID NO: 101-190, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the polypeptide is encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-100, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Antibody, Agonist and Antagonist Administration

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an antibody (including fragment or derivative thereof) of the present invention is administered. As is well-known, antibody compositions are administered, for example, to antagonize activity of CSP, or to target therapeutic agents to sites of CSP presence and/or accumulation. In a preferred embodiment, the antibody specifically binds to a CSP comprising an amino acid sequence of SEQ ID NO: 101-190, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antibody specifically binds to a CSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-100, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

The present invention also provides methods for identifying modulators which bind to a CSP or have a modulatory effect on the expression or activity of a CSP. Modulators which decrease the expression or activity of CSP (antagonists) are believed to be useful in treating colon cancer. Such screening assays are known to those of skill in the art and include, without limitation, cell-based assays and cell-free assays. Small molecules predicted via computer imaging to specifically bind to regions of a CSP can also be designed, synthesized and tested for use in the imaging and treatment of colon cancer. Further, libraries of molecules can be screened for potential anticancer agents by assessing the ability of the molecule to bind to the CSPs identified herein. Molecules identified in the library as being capable of binding to a CSP are key candidates for further evaluation for use in the treatment of colon cancer. In a preferred embodiment, these molecules will downregulate expression and/or activity of a CSP in cells.

In another embodiment of the therapeutic methods of the present invention, a pharmaceutical composition comprising a non-antibody antagonist of CSP is administered.

Antagonists of CSP can be produced using methods generally known in the art. In particular, purified CSP can be used to screen libraries of pharmaceutical agents, often combinatorial libraries of small molecules, to identify those that specifically bind and antagonize at least one activity of a CSP.

- 5 In other embodiments a pharmaceutical composition comprising an agonist of a CSP is administered. Agonists can be identified using methods analogous to those used to identify antagonists.

 In a preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a CSP comprising an amino acid sequence of SEQ
10 ID NO: 101-190, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a CSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1-100, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

15

Targeting Colon Tissue

 The invention also provides a method in which a polypeptide of the invention, or an antibody thereto, is linked to a therapeutic agent such that it can be delivered to the colon or to specific cells in the colon. In a preferred embodiment, an anti-CSP antibody is
20 linked to a therapeutic agent and is administered to a patient in need of such therapeutic agent. The therapeutic agent may be a toxin, if colon tissue needs to be selectively destroyed. This would be useful for targeting and killing colon cancer cells. In another embodiment, the therapeutic agent may be a growth or differentiation factor, which would be useful for promoting colon cell function.

25 In another embodiment, an anti-CSP antibody may be linked to an imaging agent that can be detected using, *e.g.*, magnetic resonance imaging, CT or PET. This would be useful for determining and monitoring colon function, identifying colon cancer tumors, and identifying noncancerous colon diseases.

EXAMPLES**Example 1: Gene Expression analysis**

Identification of CSGs was carried out by a systematic analysis of gene expression data in the LIFESEQ® Gold database available from Incyte Genomics Inc, Palo Alto, CA, using the data mining software package CLASP™.

The CLASP target gene identification process is focused on, but not limited to, the following 5 CLASP profiles: tissue specific expression, cancer specific expression, differentially expressed in cancer, maximum tissue differential expression.

(1) For these profiles: cDNA libraries were divided into 48 unique tissue organs. The genes were grouped into gene bins, each bin is a sequence based cluster grouped together with a common contig. The expression levels for each gene bin were calculated in each organ. Differential expression significance was calculated with rigorous statistical significant test considering the influence of sequence random fluctuations and sampling size of cDNA libraries from concept published by Audic S and Claverie JM (Genome Res 1997 7(10): 986-995: The significance of digital gene expression profiles).

(2) Highly expressed organ specific genes were selected based on the percentage abundance level in the targeted organ versus all the other organs (organ-specificity).

(3) The expression levels of each highly expressed organ-specific gene in the tumor tissue libraries were compared with normal tissue libraries and tissue libraries associated with tumor or disease (cancer-specificity) and analyzed for statistical significance.

(4) Target genes exhibiting each CLASP profile criteria were selected

CLASP 1 tissue specific expression profile: In order to meet the organ-specificity criteria, the expression level of the component clones which the gene is composed of must exhibit 3 or more occurrences regardless the total number of genes isolated for the target organ. The percentage abundance level in each organ was calculated to identify the organ with the highest expression percentage level.

CLASP 2 cancer specific expression profile: In order to fulfill the cancer specific criteria, genes must exhibit 0 expression in normal and libraries associated with tumor and

disease but not tumor per se. If the gene then exhibited organ-specificity, the gene was selected as a CLASP target for this profile.

CLASP 3 maximum tissue differential expression profile: CLASP targets were selected based on ratio of expression in tumor libraries compared to expression in normal libraries (including normal libraries associated with tumor or disease) for each organ regardless of whether the gene exhibited organ-specificity. This profile was divided into 2 sub-profiles, since the ratio of expression cannot be obtained if no expression is present in normal libraries (including normal libraries associated with tumor or disease). In this case, the maximum expression percentage of the gene, as calculated by the occurrence of the gene divided by the occurrence of all genes in the target organ, was used. CLASP selects the top 50 targets for each sub-profile.

CLASP 4 maximum tissue differential expression profile with negligible expression in normal tissues: CLASP targets were selected based on ratio of expression in tumor libraries compared to expression in normal libraries (including normal libraries associated with tumor or disease) for each organ regardless of whether the gene exhibited organ-specificity.

CLASP 5 differentially expressed in cancer profile: Expression levels in tumor libraries in each organ and normal libraries (including normal libraries associated with cancer or disease) for all organs were obtained and statistically analyzed. If the gene exhibited 90% of confidence that it is over-expressed in tumor libraries in the target organ than normal libraries for all organs, it was selected as a CLASP target for this profile. Accordingly, CLASP allows the identification of highly expressed organ and cancer specific genes based on the gene expression levels in each tissue organ. CLASP scores for a portion of the CSG of this invention are listed below.

The CLASP scores for SEQ ID NO: 1-100 are listed below:

| SEQ ID | Organ | Abundance - Percentage | Organ | Abundance - Percentage | Organ | Abundance - Percentage | Organ | Abundance - Percentage |
|--------|-------|------------------------|-------|------------------------|-------|------------------------|-------|------------------------|
| NO: 1 | INS | 3 - .0028 | | | | | | |
| NO: 2 | INS | 3 - .0028 | FTS | 2 - .0003 | | | | |
| NO: 3 | INS | 3 - .0028 | SKN | 2 - .003 | EYE | 1 - .0062 | | |
| NO: 4 | INS | 3 - .0028 | UTR | 1 - .0004 | OVR | 1 - .0007 | | |
| NO: 5 | INS | 3 - .0028 | INL | 1 - .0004 | LNG | 1 - .0003 | TST | 1 - .0011 |
| NO: 6 | INS | 3 - .0028 | BRN | 1 - .0001 | THR | 1 - .002 | | |
| NO: 7 | INS | 3 - .0028 | | | | | | |
| NO: | INS | 2 - .0257 | | | | | | |

| | | | | | | | | |
|-----------|-----|-----------|-----|-----------|-----|-----------|--|--|
| 8 | | | | | | | | |
| NO: 9 | INL | 5 - .002 | INS | 3 - .0028 | THY | 1 - .0019 | | |
| NO: 10 | INS | 5 - .0047 | INL | 1 - .0004 | | | | |
| NO: 11 | INL | 5 - .002 | TNS | 3 - .0049 | UNC | 1 - .0011 | | |
| NO: 12 | INL | 4 - .0016 | | | | | | |
| NO: 13 | INL | 3 - .0012 | | | | | | |
| NO: 14 | INL | 4 - .0016 | | | | | | |
| NO: 15 | INS | 5 - .0047 | INL | 5 - .002 | PAN | 1 - .0008 | | |
| NO: 16 | INS | 5 - .0047 | INL | 5 - .002 | | | | |
| NO: 17 | INL | 3 - .0012 | | | | | | |
| NO: 18 | INL | 3 - .0012 | | | | | | |
| NO: 19 | INL | 3 - .0012 | | | | | | |
| NO: 20 | INL | 3 - .0012 | PRO | 1 - .0003 | | | | |
| NO: 21 | INL | 3 - .0012 | | | | | | |
| NO: 22 | INL | 2 - .0043 | | | | | | |
| NO: 23 | INS | 3 - .0028 | | | | | | |
| NO: 24 | INL | 8 - .0032 | | | | | | |
| NO: 25 | INL | 8 - .0032 | | | | | | |
| NO: 26 | INL | 8 - .0032 | | | | | | |
| NO: 27 | INL | 8 - .0032 | | | | | | |
| NO: 28 | INL | 3 - .0012 | | | | | | |
| NO: 29 | INL | 2 - .0047 | | | | | | |
| NO: 30 | INL | 9 - .0036 | INS | 1 - .0009 | GEM | 2 - .0042 | | |
| NO: 31 | INL | 9 - .0036 | INS | 1 - .0009 | | | | |
| NO: 32 | INL | 2 - .0059 | | | | | | |
| NO: 33 | INL | 3 - .0012 | | | | | | |
| NO: 34 | INS | 3 - .0028 | MSL | 1 - .002 | BLV | 1 - .0006 | | |
| NO: 35 | INS | 3 - .0028 | | | | | | |
| NO: 36 | INS | 2 - .0162 | | | | | | |
| NO: 37 | INS | 4 - .0038 | | | | | | |
| NO: 38 | INL | 3 - .0012 | | | | | | |
| NO: 39 | INS | 2 - .0353 | | | | | | |
| NO: 40 | INS | 2 - .0019 | | | | | | |
| NO: | INS | 2 - .0019 | | | | | | |

| | | | | | | | | |
|-----|-----|------------|-----|-----------|-----|-----------|-----|----------|
| 41 | | | | | | | | |
| NO: | INS | 2 - .0019 | | | | | | |
| 42 | | | | | | | | |
| NO: | INL | 2 - .0052 | MAM | 1 - .0007 | | | | |
| 43 | | | | | | | | |
| NO: | INL | 3 - .0012 | STO | 1 - .0021 | | | | |
| 44 | | | | | | | | |
| NO: | INL | 3 - .0012 | | | | | | |
| 45 | | | | | | | | |
| NO: | INL | 3 - .0012 | ADR | 1 - .0013 | | | | |
| 46 | | | | | | | | |
| NO: | INS | 3 - .0028 | NRV | 1 - .0008 | | | | |
| 47 | | | | | | | | |
| NO: | INS | 2 - .0019 | | | | | | |
| 48 | | | | | | | | |
| NO: | INS | 3 - .0028 | | | | | | |
| 49 | | | | | | | | |
| NO: | INS | 3 - .0028 | BON | 1 - .002 | LMN | 1 - .0017 | | |
| 50 | | | | | | | | |
| NO: | INS | 3 - .0028 | | | | | | |
| 51 | | | | | | | | |
| NO: | INS | 4 - .0038 | INL | 2 - .0008 | | | | |
| 52 | | | | | | | | |
| NO: | INL | 3 - .0012 | KID | 2 - .0012 | | | | |
| 53 | | | | | | | | |
| NO: | INL | 3 - .0012 | | | | | | |
| 54 | | | | | | | | |
| NO: | INL | 4 - .0016 | | | | | | |
| 55 | | | | | | | | |
| NO: | INS | 5 - .0047 | CON | 1 - .0007 | | | | |
| 56 | | | | | | | | |
| NO: | INL | 3 - .0012 | | | | | | |
| 57 | | | | | | | | |
| NO: | INS | 2 - .0019 | INL | 1 - .0006 | LIV | 2 - .0038 | BLO | 1 - .004 |
| 58 | | | | | | | | |
| NO: | INL | 3 - .0012 | | | | | | |
| 59 | | | | | | | | |
| NO: | INL | 3 - .0012 | INS | 2 - .0019 | | | | |
| 60 | | | | | | | | |
| NO: | INL | 3 - .0012 | PNS | 1 - .0022 | | | | |
| 61 | | | | | | | | |
| NO: | INL | 3 - .0012 | | | | | | |
| 62 | | | | | | | | |
| NO: | INL | 3 - .0012 | | | | | | |
| 63 | | | | | | | | |
| NO: | INS | 3 - .0028 | | | | | | |
| 64 | | | | | | | | |
| NO: | INL | 3 - .0012 | | | | | | |
| 65 | | | | | | | | |
| NO: | INS | 3 - .0028 | | | | | | |
| 66 | | | | | | | | |
| NO: | INL | 3 - .0012 | | | | | | |
| 67 | | | | | | | | |
| NO: | INS | 3 - .0028 | | | | | | |
| 68 | | | | | | | | |
| NO: | INS | 3 - .0028 | | | | | | |
| 69 | | | | | | | | |
| NO: | INS | 2 - .0019 | INL | 1 - .0006 | | | | |
| 70 | | | | | | | | |
| NO: | INS | 3 - .0028 | | | | | | |
| 71 | | | | | | | | |
| NO: | INS | 2 - .0019 | | | | | | |
| 72 | | | | | | | | |
| NO: | INL | 17 - .0067 | ESO | 1 - .0039 | | | | |
| 73 | | | | | | | | |
| NO: | INL | 17 - | | | | | | |

| | | | | | | | | |
|-----|-----|------------|-----|-----------|-----|-----------|--|--|
| 74 | | .0067 | | | | | | |
| NO: | INS | 3 - .0028 | | | | | | |
| 75 | | | | | | | | |
| NO: | INS | 3 - .0028 | INL | 3 - .0012 | PIT | 1 - .0029 | | |
| 76 | | | | | | | | |
| NO: | INS | 2 - .0019 | | | | | | |
| 77 | | | | | | | | |
| NO: | INS | 2 - .0019 | | | | | | |
| 78 | | | | | | | | |
| NO: | INS | 2 - .0019 | | | | | | |
| 79 | | | | | | | | |
| NO: | INS | 2 - .0019 | | | | | | |
| 80 | | | | | | | | |
| NO: | INS | 2 - .0353 | | | | | | |
| 81 | | | | | | | | |
| NO: | INS | 2 - .0353 | | | | | | |
| 82 | | | | | | | | |
| NO: | INS | 2 - .0353 | | | | | | |
| 83 | | | | | | | | |
| NO: | INS | 2 - .0353 | | | | | | |
| 84 | | | | | | | | |
| NO: | INS | 2 - .0162 | | | | | | |
| 85 | | | | | | | | |
| NO: | INS | 2 - .0162 | | | | | | |
| 86 | | | | | | | | |
| NO: | INS | 3 - .0028 | | | | | | |
| 87 | | | | | | | | |
| NO: | INS | 3 - .0028 | | | | | | |
| 88 | | | | | | | | |
| NO: | INS | 3 - .0028 | | | | | | |
| 89 | | | | | | | | |
| NO: | INS | 2 - .0162 | | | | | | |
| 90 | | | | | | | | |
| NO: | INL | 6 - .0024 | INS | 1 - .0009 | PTH | 1 - .004 | | |
| 91 | | | | | | | | |
| NO: | INT | 2 - .03 | NOS | 1 - .0287 | | | | |
| 92 | | | | | | | | |
| NO: | INT | 2 - .03 | | | | | | |
| 93 | | | | | | | | |
| NO: | INS | 17 - .0161 | INL | 1 - .0004 | | | | |
| 94 | | | | | | | | |
| NO: | INS | 17 - .0161 | INL | 1 - .0004 | | | | |
| 95 | | | | | | | | |
| NO: | INS | 4 - .0038 | INL | 1 - .0004 | | | | |
| 96 | | | | | | | | |
| NO: | INS | 3 - .0028 | | | | | | |
| 97 | | | | | | | | |
| NO: | INS | 3 - .0028 | | | | | | |
| 98 | | | | | | | | |
| NO: | INS | 4 - .0038 | INL | 2 - .0008 | | | | |
| 99 | | | | | | | | |
| NO: | INS | 4 - .0038 | INL | 2 - .0008 | | | | |
| 100 | | | | | | | | |

5 INS Intestine, Small
 INT Intestine
 INL Intestine, Large
 CON Connective Tissue
 MSL Muscles
 NRV Nervous Tissue
 10 SKN Skin
 NOS Nose
 UNC Mixed Tissues
 EYE Eye

OVR Ovary
 UTR Uterus
 FTS Fetus
 KID Kidney

- 5 There are 2 values for each organ in the format 9 - 0.9999. The first represent the number of occurrences of the gene in the given organ. The 2nd number represents the percentage of the expression of the gene in the given organ.

Based on sequence alignment with the human genome, the following chromosomal locations were assigned.

| Nucleotide Name | Chromosomal Location | ORF location | Amino acid Name |
|-----------------|----------------------|--------------|-----------------|
| DEX0234_1 | 3p26.3c | 304-125 | DEX0234_101 |
| DEX0234_2 | 5p15.33e | 717-523 | DEX0234_102 |
| DEX0234_3 | 6p25.3b | 136-285 | DEX0234_103 |
| DEX0234_4 | 6p25.3b | | |
| DEX0234_5 | 2p25.3g | 123-1 | DEX0234_104 |
| DEX0234_6 | 1p36.33b | 620-775 | DEX0234_105 |
| DEX0234_7 | 1p36.33b | 160-23 | DEX0234_106 |
| DEX0234_8 | 19p13.3j | 471-316 | DEX0234_107 |
| DEX0234_9 | 12p13.33d | 1-2787 | DEX0234_108 |
| DEX0234_10 | 17p13.3g | 58-1113 | DEX0234_109 |
| DEX0234_11 | 2p25.3g | 165-2312 | DEX0234_110 |
| DEX0234_12 | 6p25.3b | 131-328 | DEX0234_111 |
| DEX0234_13 | 19p13.3j | 66-1 | DEX0234_112 |
| DEX0234_14 | 13p13e | 19-84 | DEX0234_113 |
| DEX0234_15 | 11p15.5d | 157-32 | DEX0234_114 |
| DEX0234_16 | 11p15.5d | 32-160 | DEX0234_115 |
| DEX0234_17 | 5p15.33e | | |
| DEX0234_18 | 4p16.3d | 31-117 | DEX0234_116 |
| DEX0234_19 | 8p23.3b | 349-56 | DEX0234_117 |
| DEX0234_20 | 19p13.3j | 462-872 | DEX0234_118 |
| DEX0234_21 | 19p13.3j | 231-386 | DEX0234_119 |
| DEX0234_22 | 16p13.3f | 907-638 | DEX0234_120 |
| DEX0234_23 | 12p13.33d | 713-904 | DEX0234_121 |
| DEX0234_24 | 2p25.3g | 102-1 | DEX0234_122 |
| DEX0234_25 | 2p25.3g | 596-1051 | DEX0234_123 |
| DEX0234_26 | 2p25.3g | 465-644 | DEX0234_124 |
| DEX0234_27 | 2p25.3g | 865-614 | DEX0234_125 |
| DEX0234_28 | 10p15.3d | 288-184 | DEX0234_126 |
| DEX0234_29 | | | |
| DEX0234_30 | 19p13.3j | 1117-3705 | DEX0234_127 |
| DEX0234_31 | 19p13.3j | 120-683 | DEX0234_128 |
| DEX0234_32 | 15p13e | 48-13 | DEX0234_129 |
| DEX0234_33 | 16p13.3f | 1938-1609 | DEX0234_130 |
| DEX0234_34 | 9p24.3b | 533-916 | DEX0234_131 |
| DEX0234_35 | 2p25.3g | 573-809 | DEX0234_132 |
| DEX0234_36 | 8p23.3b | 279-109 | DEX0234_133 |
| DEX0234_37 | 20p13f | 351-190 | DEX0234_134 |
| DEX0234_38 | 11p15.5d | 186-43 | DEX0234_135 |
| DEX0234_39 | 9p24.3b | 1-489 | DEX0234_136 |
| DEX0234_40 | 7p22.3d | 546-686 | DEX0234_137 |
| DEX0234_41 | 7p22.3d | 321-178 | DEX0234_138 |
| DEX0234_42 | 7p22.3d | 40-246 | DEX0234_139 |
| DEX0234_43 | 13p13e | 39-1 | DEX0234_140 |

| | | | |
|-------------|-----------|-----------|-------------|
| DEX0234_44 | 3p26.3c | 326-3 | DEX0234_141 |
| DEX0234_45 | 3p26.3c | 304-188 | DEX0234_142 |
| DEX0234_46 | 10p15.3d | 343-119 | DEX0234_143 |
| DEX0234_47 | 5p15.33e | 2153-1794 | DEX0234_144 |
| DEX0234_48 | 11p15.5d | 68-1 | DEX0234_145 |
| DEX0234_49 | 1p36.33b | 39-140 | DEX0234_146 |
| DEX0234_50 | 2p25.3g | 810-1202 | DEX0234_147 |
| DEX0234_51 | 12p13.33d | 375-220 | DEX0234_148 |
| DEX0234_52 | 7p22.3d | | |
| DEX0234_53 | 13p13e | 733-605 | DEX0234_149 |
| DEX0234_54 | 2p25.3g | 529-395 | DEX0234_150 |
| DEX0234_55 | 3p26.3c | 204-1 | DEX0234_151 |
| DEX0234_56 | 6p25.3b | 275-3 | DEX0234_152 |
| DEX0234_57 | 3p26.3c | | |
| DEX0234_58 | 16p13.3f | 107-577 | DEX0234_153 |
| DEX0234_59 | 15p13e | 320-388 | DEX0234_154 |
| DEX0234_60 | 1p36.33b | | |
| DEX0234_61 | 17p13.3g | 288-148 | DEX0234_155 |
| DEX0234_62 | 17p13.3g | 698-543 | DEX0234_156 |
| DEX0234_63 | 15p13e | 1574-1356 | DEX0234_157 |
| DEX0234_64 | 4p16.3d | 16-114 | DEX0234_158 |
| DEX0234_65 | 20p13f | 334-275 | DEX0234_159 |
| DEX0234_66 | 19p13.3j | | |
| DEX0234_67 | 17p13.3g | 137-1 | DEX0234_160 |
| DEX0234_68 | 8p23.3b | 140-1 | DEX0234_161 |
| DEX0234_69 | 17p13.3g | 459-872 | DEX0234_162 |
| DEX0234_70 | 2p25.3g | 1596-1390 | DEX0234_163 |
| DEX0234_71 | 11p15.5d | 52-2 | DEX0234_164 |
| DEX0234_72 | 3p26.3c | 466-2820 | DEX0234_165 |
| DEX0234_73 | 17p13.3g | 884-1174 | DEX0234_166 |
| DEX0234_74 | 17p13.3g | 940-293 | DEX0234_167 |
| DEX0234_75 | 17p13.3g | 257-460 | DEX0234_168 |
| DEX0234_76 | 15p13e | 2214-1906 | DEX0234_169 |
| DEX0234_77 | 6p25.3b | 461-577 | DEX0234_170 |
| DEX0234_78 | 1p36.33b | 157-2 | DEX0234_171 |
| DEX0234_79 | 7p22.3d | 134-1 | DEX0234_172 |
| DEX0234_80 | 15p13e | | |
| DEX0234_81 | 7p22.3d | 448-549 | DEX0234_173 |
| DEX0234_82 | 9p24.3b | 12-101 | DEX0234_174 |
| DEX0234_83 | 3p26.3c | 9-71 | DEX0234_175 |
| DEX0234_84 | 5p15.33e | 158-394 | DEX0234_176 |
| DEX0234_85 | 10p15.3d | 9-122 | DEX0234_177 |
| DEX0234_86 | 20p13f | 293-358 | DEX0234_178 |
| DEX0234_87 | 5p15.33e | 246-416 | DEX0234_179 |
| DEX0234_88 | 2p25.3g | | |
| DEX0234_89 | 7p22.3d | | |
| DEX0234_90 | Xp22.33f | 228-299 | DEX0234_180 |
| DEX0234_91 | 2p25.3g | 590-312 | DEX0234_181 |
| DEX0234_92 | 16p13.3f | 56-139 | DEX0234_182 |
| DEX0234_93 | 16p13.3f | 90-590 | DEX0234_183 |
| DEX0234_94 | 17p13.3g | 222-464 | DEX0234_184 |
| DEX0234_95 | 17p13.3g | 105-584 | DEX0234_185 |
| DEX0234_96 | 12p13.33d | 1-2787 | DEX0234_186 |
| DEX0234_97 | 2p25.3g | 688-398 | DEX0234_187 |
| DEX0234_98 | 10p15.3d | 376-233 | DEX0234_188 |
| DEX0234_99 | 4p16.3d | 317-2251 | DEX0234_189 |
| DEX0234_100 | 4p16.3d | 252-106 | DEX0234_190 |

Example 2A: Custom Microarray Experiment—Colon Cancer

Custom oligonucleotide microarrays were provided by Agilent Technologies, Inc. (Palo Alto, CA). The microarrays were fabricated by Agilent using their technology for the *in-situ* synthesis of 60mer oligonucleotides (Hughes, et al. 2001, Nature

5 Biotechnology 19:342-347). The 60mer microarray probes were designed by Agilent, from gene sequences provided by diaDexus, using Agilent proprietary algorithms. Whenever possible two different 60mers were designed for each gene of interest.

All microarray experiments were two-color experiments and were preformed using Agilent-recommended protocols and reagents. Briefly, each microarray was
10 hybridized with cRNAs synthesized from polyA+ RNA, isolated from cancer and normal tissues, labeled with fluorescent dyes Cyanine3 and Cyanine5 (NEN Life Science Products, Inc., Boston, MA) using a linear amplification method (Agilent). In each experiment the experimental sample was polyA+ RNA isolated from cancer tissue from a single individual and the reference sample was a pool of polyA+ RNA isolated from
15 normal tissues of the same organ as the cancerous tissue (*i.e.* normal colon tissue in experiments with colon cancer samples). Hybridizations were carried out at 60°C, overnight using Agilent *in-situ* hybridization buffer. Following washing, arrays were scanned with a GenePix 4000B Microarray Scanner (Axon Instruments, Inc., Union City, CA). The resulting images were analyzed with GenePix Pro 3.0 Microarray Acquisition
20 and Analysis Software (Axon). Two different chip designs were evaluated with overlapping sets of a total of 38 samples, comparing the expression patterns of colon cancer derived polyA+ RNA to polyA+ RNA isolated from a pool of 7 normal colon tissues were analyzed. For Chip1 all 38 samples (23 Ascending colon carcinomas and 15 Rectosigmoidal carcinomas including: 5 stage 1 cancers, 15 stage 2 cancers, 15 stage 3
25 and 2 stage 4 cancers) were analyzed and for Chip2 a subset of 27 of these samples (14 Ascending colon carcinomas and 13 Rectosigmoidal carcinomas including: 3 stage 1 cancers, 9 stage 2 cancers, 13 stage 3 and 2 stage 4 cancers) were assessed.

Data normalization and expression profiling were done with Expressionist software from GeneData Inc. (Daly City, CA/Basel, Switzerland). Gene expression
30 analysis was performed using only experiments that meet certain quality criteria. The quality criteria that experiments must meet are a combination of evaluations performed by the Expressionist software and evaluations performed manually using raw and normalized data. To evaluate raw data quality, detection limits (the mean signal for a replicated

negative control + 2 Standard Deviations (SD)) for each channel were calculated. The detection limit is a measure of non-specific hybridization. Arrays with poor detection limits were not analyzed and the experiments were repeated. To evaluate normalized data quality, positive control elements included in the array were utilized. These array features should have a mean ratio of 1 (no differential expression). If these features have a mean ratio of greater than 1.5-fold up or down, the experiments were not analyzed further and were repeated. In addition to traditional scatter plots demonstrating the distribution of signal in each experiment, the Expressionist software also has minimum thresholding criteria that employ user defined parameters to identify quality data. Only those features that meet the threshold criteria were included in the filtering and analyses carried out by Expressionist. The thresholding settings employed require a minimum area percentage of 60% [(% pixels > background + 2SD)-(% pixels saturated)], and a minimum signal to noise ratio of 2.0 in both channels. By these criteria, very low expressors and saturated features were not included in analysis.

Relative expression data was collected from Expressionist based on filtering and clustering analyses. Up- and down- regulated genes were identified using criteria for percentage of valid values obtained, and the percentage of experiments in which the gene is up- or down-regulated. These criteria were set independently for each data set, depending on the size and the nature of the data set. The results for the statistically significant upregulated and downregulated genes are shown in Table 1 and Table 2. The first three columns of each table contain information about the sequence itself (Oligo ID, Parent ID, and Patent#), the next 3 columns show the results obtained. '%valid' indicates the percentage of unique experiments total (n=38 for Chip1, n=27 for Chip2) in which a valid expression value was obtained, '%up' indicates the percentage of experiments in which up-regulation of at least 2-fold was observed, and '%down' indicates the percentage of the experiments in which down-regulation of at least 2-fold was observed. The last column in each table describes the location of the microarray probe (oligo) relative to the parent sequence.

Table 1. Colon Microarray expression data.

| Patent | ParentID | OligoID | %valid n=38 | %down n=38 | %valid down n=38 | Oligo Start Pos in Par Seq | Oligo Stop Pos in Par Seq |
|------------|----------|---------|----------------|---------------|------------------------|-------------------------------------|------------------------------------|
| DEX0234_11 | 11855 | 26273 | 94.7 | 68.4 | 72.2 | 158 | 217 |
| DEX0234_11 | 11855 | 26272 | 92.1 | 60.5 | 65.7 | 209 | 268 |
| DEX0234_15 | 11865 | 23559 | 13.2 | 2.6 | 20 | 201 | 260 |

| Patent | ParentID | OligoID | %valid n=38 | %down n=38 | %valid down n=38 | Oligo Start Pos in Par Seq | Oligo Stop Pos in Par Seq |
|---------------------------|----------|---------|----------------|---------------|------------------------|-------------------------------------|------------------------------------|
| DEX0234_16 | 11865 | 23558 | 86.8 | 86.8 | 100 | 611 | 670 |
| DEX0234_27, DEX0234_24 | 11897 | 38407 | 100 | 68.4 | 68.4 | 1092 | 1151 |
| DEX0234_27 | 11897 | 38408 | 100 | 63.2 | 63.2 | 1051 | 1110 |
| DEX0234_52 | 11696 | 27384 | 94.7 | 63.2 | 66.7 | 968 | 1027 |
| DEX0234_52 | 11696 | 27383 | 100 | 63.2 | 63.2 | 1028 | 1087 |
| DEX0234_60 | 11848 | 34228 | 65.8 | 55.3 | 84 | 1642 | 1701 |
| DEX0234_60 | 11848 | 34229 | 52.6 | 28.9 | 55 | 1245 | 1304 |

Table 2. Colon Microarray Tissue Descriptions.

| Experiment Name | Renamed | Tissue ID | Region affected | Stage | Grade | TNM |
|--------------------|-------------|------------|--------------------|-------|--------|---------|
| CNA1C038NMP | | BS993038A | | 1 | 1 | T1N0M0 |
| CNA1C089NMP | | S9914089A | | 1 | 1 or 2 | T1N0M0 |
| CNA1C108NMP | | 9706C108RA | | 1 | 2 | T2N0M0 |
| CNA1C281NMP | | S992281A | | 1 | 1 | T2N0M0 |
| CNA1C795NMP | | S9915795A | | 1 | 1 or 2 | T1N0M0 |
| CNA2C012NMP | | 9608B012 | | 2 | 2-3 | T4N0M0 |
| CNA2C062NMP | | S993062C | | 2 | 1 or 2 | T3N0M0 |
| CNA2C074NMP | | 9709C074RA | | 2 | 2-3 | T3N0M0 |
| CNA2C382NMP | | 4005382A3 | | 2 | 2 | T2N0M0 |
| CNA2C56CNMP | | 1056C | | 2 | 2 | T3N0M0 |
| CNA2C677NMP | | 4005677A1 | | 2 | 3 | T3N0M0 |
| CNA2C695NMP | | 4004695A9 | | 2 | 2 | T3N0M0 |
| CNA2C821NMP | | S9914821A | | 2 | 2 | T3N0M0 |
| CNA3C003NMP | | 9610B003 | | 3 | 3 | T3M1M0 |
| CNA3C004NMP | | 9707C004GB | | 3 | 2 | T3N1M0 |
| CNA3C032NMP | | S9921032A | | 3 | 2 | T3N1M0 |
| CNA3C068NMP | | 9706C068RA | | 3 | 2 or 3 | T3N1M0 |
| CNA3C401NMP | | S9819401A | | 3 | 2 | T2N1M0 |
| CNA3C720NMP | | S993720A | | 3 | 3 | T1N2M0 |
| CNA3C806NMP | CNA4C806NMP | S9915806A | | 4 | 2 | T3N1M1 |
| CNA3C810NMP | | BS986810A | | 3 | 2 | T41N1M0 |
| CNA4C005NMP | | 9706C005RA | | 4 | 3 | T3N2M1 |
| CNA4C006NMP | | 9609B006 | | 4 | 2 or 3 | T3N2M1 |
| CNR2C020NMP | | 9408C020R | | 2 | 2 | T3N0M0 |
| CNR2C024NMP | | 9704C024RA | | 2 | 2 | T3N0M0 |
| CNR2C036ANMP | | 9705C036A | | 2 | 2 | T3N0M0 |
| CNR2C086NMP | | 9707C086B | | 2 | 2 | T3N0M0 |
| CNR2C162NMP | | 9406C162R | | 2 | 2 | T4N0M0 |
| CNR2C196NMP | | S9820196A | | 2 | 2 | T3N0M0 |
| CNR2C404NMP | | S9819404A | | 2 | 2 | T4N0M0 |
| CNR3C006BNMP | | 9702B006B | | 3 | 2 | T3N1M0 |
| CNR3C014NMP | | 9707C014RA | | 3 | 2 | T3N2M0 |
| CNR3C022NMP | | 9611B022F | | 3 | 2 | T3N1M0 |
| CNR3C036NMP | | 9706C036RA | | 3 | 2 | T3N2M0 |
| CNR3C053NMP | | 9409C053R | | 3 | 2 | T3N1M0 |
| CNR3C091NMP | CNR4C091NMP | 9703C091R | | 4 | 3 | T3N1M1 |
| CNR3C457NMP | | 4004457A3 | | 3 | 3 | T3N1M0 |
| CNR3CC98NMP | | 1194C98 | | 3 | 2 | T3N1M0 |

Example 2B :Relative Quantitation of Gene Expression

Real-Time quantitative PCR with fluorescent Taqman probes is a quantitation detection system utilizing the 5'-3' nuclease activity of Taq DNA polymerase. The method uses an internal fluorescent oligonucleotide probe (Taqman) labeled with a 5' reporter dye and a downstream, 3' quencher dye. During PCR, the 5'-3' nuclease activity of Taq DNA polymerase releases the reporter, whose fluorescence can then be detected by the laser detector of the Model 7700 Sequence Detection System (PE Applied Biosystems, Foster City, CA, USA). Amplification of an endogenous control is used to standardize the amount of sample RNA added to the reaction and normalize for Reverse Transcriptase (RT) efficiency. Either cyclophilin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), ATPase, or 18S ribosomal RNA (rRNA) is used as this endogenous control. To calculate relative quantitation between all the samples studied, the target RNA levels for one sample were used as the basis for comparative results (calibrator). Quantitation relative to the "calibrator" can be obtained using the standard curve method or the comparative method (User Bulletin #2: ABI PRISM 7700 Sequence Detection System).

The tissue distribution and the level of the target gene are evaluated for every sample in normal and cancer tissues. Total RNA is extracted from normal tissues, cancer tissues, and from cancers and the corresponding matched adjacent tissues. Subsequently, first strand cDNA is prepared with reverse transcriptase and the polymerase chain reaction is done using primers and Taqman probes specific to each target gene. The results are analyzed using the ABI PRISM 7700 Sequence Detector. The absolute numbers are relative levels of expression of the target gene in a particular tissue compared to the calibrator tissue.

One of ordinary skill can design appropriate primers. The relative levels of expression of the CSNA versus normal tissues and other cancer tissues can then be determined. All the values are compared to normal thymus (calibrator). These RNA samples are commercially available pools, originated by pooling samples of a particular tissue from different individuals.

The relative levels of expression of the CSNA in pairs of matching samples and 1 cancer and 1 normal/normal adjacent of tissue may also be determined. All the values are compared to normal thymus (calibrator). A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

In the analysis of matching samples, the CSNAs that show a high degree of tissue specificity for the tissue of interest. These results confirm the tissue specificity results obtained with normal pooled samples. Further, the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual are compared.

- 5 This comparison provides an indication of specificity for the cancer stage (*e.g.* higher levels of mRNA expression in the cancer sample compared to the normal adjacent).

Altogether, the high level of tissue specificity, plus the mRNA overexpression in matching samples tested are indicative of SEQ ID NO: 1-100 being a diagnostic marker for cancer.

10 **Example 3: Protein Expression**

- The CSNA is amplified by polymerase chain reaction (PCR) and the amplified DNA fragment encoding the CSNA is subcloned in pET-21d for expression in *E. coli*. In addition to the CSNA coding sequence, codons for two amino acids, Met-Ala, flanking the NH₂-terminus of the coding sequence of CSNA, and six histidines, flanking the
- 15 COOH-terminus of the coding sequence of CSNA, are incorporated to serve as initiating Met/restriction site and purification tag, respectively.

An over-expressed protein band of the appropriate molecular weight may be observed on a Coomassie blue stained polyacrylamide gel. This protein band is confirmed by Western blot analysis using monoclonal antibody against 6X Histidine tag.

- 20 Large-scale purification of CSP is achieved using cell paste generated from 6-liter bacterial cultures, and purified using immobilized metal affinity chromatography (IMAC). Soluble fractions that are separated from total cell lysate were incubated with a nickle chelating resin. The column is packed and washed with five column volumes of wash buffer. CSP is eluted stepwise with various concentration imidazole buffers.

25 **Example 4: Fusion Proteins**

- The human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector. For example, if pC4 (Accession No.
- 30 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the

present invention, isolated by the PCR protocol described in Example 2, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced. If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if
5 the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. *See, e. g.*, WO 96/34891.

Example 5: Production of an Antibody from a Polypeptide

In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such
10 cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100, µg/ml of streptomycin. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any
15 suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP20), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands *et al.*, *Gastroenterology* 80: 225-232 (1981).

20 The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide. Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody
25 which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic
30 antibodies to the protein specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

For the polypeptides of the invention, the following attributes were found, epitopes, post translational modifications, signal peptides and transmembrane domains. Specifically, the Jameson-Wolf methods were used to predict epitopes were predicted. (Jameson and Wolf, CABIOS, 4(1), 181-186, 1988). Examples of post-translational

5 modifications (PTMs) and other motifs of the CSPs of this invention are listed below. In addition, antibodies that specifically bind such post-translational modifications may be useful as a diagnostic or as therapeutic. The PTMs and other motifs were predicted by using the ProSite Dictionary of Proteins Sites and Patterns (Bairoch et al., Nucleic Acids Res. 25(1):217-221 (1997)), the following motifs, including PTMs, were predicted for the

10 CSPs of the invention. The signal peptides were detected by using the SignalP 2.0, see "Machine learning approaches to the prediction of signal peptides and other protein sorting signals" Henrik Nielsen, Søren Brunak, and Gunnar von Heijne, Protein Engineering 12, 3-9 (1999). Prediction of transmembrane helices in proteins was performed by the application TMHMM 2.0, "currently the best performing transmembrane

15 prediction program", according to authors (A. Krogh, B. Larsson, G. von Heijne, and E. L. Sonnhammer. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. Journal of Molecular Biology, 305(3):567-580, January 2001). S. Moller, M.D.R. Croning, R. Apweiler "Evaluation of methods for the prediction of membrane spanning regions" Bioinformatics, 17(7):646-653, July 2001. E.

20 L.L. Sonnhammer, G. von Heijne, and A. Krogh. A hidden Markov model for predicting transmembrane helices in protein sequences in J. Glasgow, T. Littlejohn, F. Major, R. Lathrop, D. Sankoff, and C. Sensen, editors, Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology, pages 175-182, Menlo Park, CA, 1998. AAAI Press. Using the PSORT II program, the following cellular localizations

25 and the k nearest neighbors classifier values were determined (Paul Horton and Kenta Nakai, Better Prediction of Protein Cellular Localization Sites with the k Nearest Neighbors Classifier, Intelligent Systems for Molecular Biology 5 147-152 (1997). In the table below are the following: PTM and other motifs (type, amino acid residue locations); Amino acid location and antigenic index (location, AI score, length); TM (number of

30 membrane domain, topology in orientation and position).

| name | DEX0234 .aa.ai | DEX0234.a a.pst | DEX0234.aa.ptm | DEX0234 .aa.sp | DEX0234. aa.tmhm |
|-------------|--|--------------------|---|-------------------|---------------------|
| DEX0234_101 | | nuc | Ck2_Phospho_Site 27-30;32-35; Cytochrome_C 36-41; Pkc_Phospho_Site 27-29; | 28 0.945 0.628 | |
| DEX0234_102 | | cyt | Asn_Glycosylation 36-39; Pkc_Phospho_Site 16-18; | | |
| DEX0234_103 | | nuc | | | |
| DEX0234_104 | | nuc | | | |
| DEX0234_105 | | pla | | | |
| DEX0234_106 | | pla | Myristyl 18-23;31-36; | | 1 i12-34o |
| DEX0234_107 | | cyt | Myristyl 14-19;38-43; Pkc_Phospho_Site 20-22;30-32; | | |
| DEX0234_108 | 53-65 1.23 13 744-755 1.19 12 864-880 1.11 17 564-581 1.06 18 588-604 1.06 17 435-460 1.03 26 697-709 1.02 13 271-281 1.02 11 | cyt | Amidation 699-702; Asn_Glycosylation 83-86;120- 123;161-164;198-201;239-242;276- 279;373-376;902-905; Atp_Gtp_A 527-534; Camp_Phospho_Site 701- 704;876-879; Ck2_Phospho_Site 11-14;70-73;91-94;169-172;247- 250;286-289;376-379;525-528;609- 612;791-794;809-812;842-845; Myristyl 127-132;205-210;282- 287;456-461;471-476;625-630;749- 754;786-791;866-871; Pkc_Phospho_Site 84-86;130- 132;162-164;208-210;240-242;414- 416;540-542;681-683;754-756;770- 772;791-793;835-837;844-846;848- 850; Transferrin_1 146-154;224- 232; | | |
| DEX0234_109 | 335-344 1.06 10 267-277 1.05 11 | ves | Asn_Glycosylation 228-231; Camp_Phospho_Site 125-128; Ck2_Phospho_Site 60-63; Leucine_Zipper 205-226; Myristyl 66-71;112-117;163-168;346-351; Pkc_Phospho_Site 109-111;124- 126;169-171; | | |

| | | | | | |
|-------------|--|-----|--|-------------------|-----------------|
| DEX0234_110 | 424-434 1.22 11 | pla | Amidation 373-376; Asn_Glycosylation 103-106;157-160;174-177;264-267;452-455;498-501;539-542;551-554;594-597; Ck2_Phospho_Site 40-43;64-67;85-88;265-268;287-290;339-342;403-406;481-484;507-510; Myristyl 144-149;354-359;383-388;419-424;511-516;544-549; Pkc_Phospho_Site 54-56;87-89;108-110;180-182;224-226;355-357;386-388;474-476;507-509;515-517;541-543;599-601;608-610;712-714; Tnfr_Ngfr_1 489-522; Tyr_Phospho_Site 139-147;292-298; Zinc_Protease 402-411; | 17 0.956 0.866 | 1 o673-695i |
| | 85-94 1.14 10 495-536 1.07 42 354-364 1.06 11 548-566 1.06 19 382-392 1.05 11 18-36 1.03 19 | | | | |
| DEX0234_111 | | nuc | Amidation 11-14; Ck2_Phospho_Site 49-52; Pkc_Phospho_Site 11-13; | | |
| DEX0234_112 | | cyt | Myristyl 12-17; | | |
| DEX0234_113 | | cyt | Tyr_Phospho_Site 12-19; | | |
| DEX0234_114 | 8-28 1.00 21 | nuc | Asn_Glycosylation 35-38; Ck2_Phospho_Site 10-13;12-15; Myristyl 28-33; Pkc_Phospho_Site 7-9;12-14; | | |
| DEX0234_115 | | nuc | Myristyl 27-32; | | |
| DEX0234_116 | | cyt | Ck2_Phospho_Site 19-22; | | |
| DEX0234_117 | | pla | Asn_Glycosylation 61-64; Myristyl 7-12; Pkc_Phospho_Site 40-42; | | 2 i13-35o64-86i |
| DEX0234_118 | 18-35 1.17 18 | exc | Camp_Phospho_Site 20-23; Ck2_Phospho_Site 64-67; Myristyl 4-9;37-42; Pkc_Phospho_Site 27-29; | | |
| DEX0234_119 | | cyt | Myristyl 43-48; | | |
| DEX0234_120 | | exc | Ck2_Phospho_Site 62-65; Pkc_Phospho_Site 83-85; | 15 0.953 0.880 | |
| DEX0234_121 | | nuc | Camp_Phospho_Site 45-48; Myristyl 47-52; Pkc_Phospho_Site 55-57; | | |
| DEX0234_122 | | nuc | Asn_Glycosylation 30-33; | | |
| DEX0234_123 | 117-137 1.10 21 | nuc | Ck2_Phospho_Site 14-17;36-39; Myristyl 46-51;113-118; Pkc_Phospho_Site 11-13;94-96;100-102;132-134; | | |
| DEX0234_124 | | pla | Myristyl 20-25; Pkc_Phospho_Site 36-38; Prokar_Lipoprotein 15-25; | 23 0.996 0.908 | |

[illegible]

| | | | | | |
|-------------|-------------------|-----|--|-------------------|-----------|
| DEX0234_133 | | mit | Myristyl 27-32;36-41;40-45; Pkc_Phospho_Site 48-50; | | 1 o22-44i |
| DEX0234_134 | | cyt | Myristyl 9-14;13-18; Pkc_Phospho_Site 14-16; | | |
| DEX0234_135 | | cyt | Myristyl 21-26; | | |
| DEX0234_136 | 51-76 1.06 26 | ves | Amidation 64-67; Myristyl 38-43;74- 79;135-140; Pkc_Phospho_Site 8- 10;52-54;154-156; | | |
| DEX0234_137 | 19-35 1.16 17 | exc | Asn_Glycosylation 43-46; Myristyl 26-31; | 15 0.992 0.951 | |
| DEX0234_138 | 18-36 1.08 19 | cyt | Asn_Glycosylation 29-32; | | |
| DEX0234_139 | 21-51 1.16 31 | nuc | Ck2_Phospho_Site 28-31;42-45; Myristyl 13-18;38-43;50-55; Pkc_Phospho_Site 35-37;41-43;42- 44; | | |
| DEX0234_141 | 49-68 1.17 20 | nuc | Asn_Glycosylation 65-68; Ck2_Phospho_Site 56-59; Pkc_Phospho_Site 25-27;30-32;56- 58;84-86; | | |
| DEX0234_142 | | cyt | Pkc_Phospho_Site 7-9; | | |
| DEX0234_143 | 41-61 1.16 21 | cyt | Ck2_Phospho_Site 68-71; Myristyl 42-47;46-51;55-60;59-64;63-68;64- 69; | | |
| DEX0234_144 | | pla | Ck2_Phospho_Site 45-48; Myristyl 5-10; Pkc_Phospho_Site 54-56;82- 84;91-93; | 52 0.996 0.764 | 1 i13-35o |
| DEX0234_145 | | nuc | Ck2_Phospho_Site 13-16; Myristyl 4-9; Pkc_Phospho_Site 8-10;20-22; | | |
| DEX0234_146 | | cyt | Ck2_Phospho_Site 8-11; Pkc_Phospho_Site 28-30; | | |
| DEX0234_147 | 96-106 1.34 11 | nuc | Ck2_Phospho_Site 77-80; Myristyl 26-31; Pkc_Phospho_Site 11-13; | | |
| DEX0234_148 | 16-26 1.04 11 | nuc | Ck2_Phospho_Site 2-5; Pkc_Phospho_Site 2-4;34-36; | | |
| DEX0234_149 | 26-40 1.16 15 | nuc | | | |
| DEX0234_150 | | cyt | Pkc_Phospho_Site 2-4; | | |
| DEX0234_151 | 34-58 1.14 25 | exc | Ck2_Phospho_Site 42-45;50-53; Pkc_Phospho_Site 42-44; | 38 0.962 0.620 | |
| DEX0234_152 | 13-24 1.19 12 | nuc | Asn_Glycosylation 22-25; Ck2_Phospho_Site 10-13;33-36; Pkc_Phospho_Site 6-8; Tyr_Phospho_Site 76-83; | | |
| DEX0234_153 | 96-136 1.03 41 | mit | Amidation 111-114; Ck2_Phospho_Site 41-44; Glycosaminoglycan 56-59;60-63; Myristyl 17-22;19-24;21-26;36- 41;39-44;57-62;59-64;61-66;91- 96;117-122;118-123;122-127;130- 135;145-150; Pkc_Phospho_Site 25- 27; Prokar_Lipoprotein 14-24; | 24 0.924 0.693 | |
| DEX0234_154 | | cyt | Ck2_Phospho_Site 17-20; | | |

| | | | | | |
|-------------|--|-----|---|--|------------|
| DEX0234_155 | | cyt | Asn_Glycosylation 13-16; Ck2_Phospho_Site 15-18; | | |
| DEX0234_156 | | cyt | Pkc_Phospho_Site 35-37; | | |
| DEX0234_157 | | nuc | Asn_Glycosylation 25-28; Ck2_Phospho_Site 66-69; Pkc_Phospho_Site 11-13;45-47; Prokar_Lipoprotein 49-59; | | |
| DEX0234_158 | | cyt | | | |
| DEX0234_160 | 9-18 1.05 10 | nuc | Ck2_Phospho_Site 37-40; Prokar_Lipoprotein 17-27; | | |
| DEX0234_161 | | nuc | Ck2_Phospho_Site 8-11; Pkc_Phospho_Site 21-23; | | |
| DEX0234_162 | | nuc | Amidation 15-18; Asn_Glycosylation 73-76; Camp_Phospho_Site 61-64; Myristyl 30-35;92-97;123-128; Pkc_Phospho_Site 64-66; | | |
| DEX0234_163 | | cyt | Amidation 62-65; Myristyl 13-18; | | |
| DEX0234_164 | | | Asn_Glycosylation 2-5; Pkc_Phospho_Site 5-7; | | |
| DEX0234_165 | 241-251 1.20 11 164-181 1.17 18 341-354 1.16 14 202-231 1.13 30 720-735 1.10 16 679-702 1.09 24 447-459 1.02 13 | nuc | Asn_Glycosylation 465-468; Camp_Phospho_Site 227-230; Ck2_Phospho_Site 3-6;31-34;38- 41;132-135;137-140;239-242;484- 487;635-638;691-694;726-729; Myristyl 46-51;97-102;167-172;183- 188;225-230;229-234;449-454; Pkc_Phospho_Site 141-143;171- 173;226-228;333-335;407-409;429- 431;501-503;508-510;654-656; Zinc_Finger_C2h2 16-37; | | |
| DEX0234_166 | 58-75 1.22 18 | cyt | Ck2_Phospho_Site 21-24; Myristyl 58-63; Pkc_Phospho_Site 26-28; | | |
| DEX0234_167 | 115-131 1.14 17 54-80 1.09 27 | nuc | Amidation 206-209; Ck2_Phospho_Site 26-29; Myristyl 16-21;20-25;47-52;105-110;130- 135;155-160;158-163; Pkc_Phospho_Site 17-19;64- 66;114-116;124-126;125-127;162- 164; | | 1 092-114i |

| | | | | | |
|-------------|--|-----|---|-------------------|-----------|
| DEX0234_168 | | nuc | Myristyl 8-13; Pkc_Phospho_Site 2-4;13-15;54-56; | | |
| DEX0234_169 | | nuc | Amidation 90-93; Myristyl 40-45;75-80;78-83; | | |
| DEX0234_170 | 11-28 1.08 18 | nuc | Amidation 19-22; Asn_Glycosylation 17-20; Ck2_Phospho_Site 6-9; Pkc_Phospho_Site 19-21; | | |
| DEX0234_171 | | nuc | Asn_Glycosylation 22-25; | | |
| DEX0234_172 | | exc | Myristyl 21-26; Pkc_Phospho_Site 9-11;31-33; | 25 0.986 0.919 | 1 i13-35o |
| DEX0234_173 | | cyt | Pkc_Phospho_Site 28-30; | | |
| DEX0234_174 | 13-25 1.18 13 | cyt | Pkc_Phospho_Site 17-19; | | |
| DEX0234_175 | | exc | | | |
| DEX0234_176 | | nuc | Camp_Phospho_Site 30-33; | | |
| DEX0234_177 | 22-34 1.03 13 | cyt | Pkc_Phospho_Site 2-4; | | |
| DEX0234_178 | | nuc | | | |
| DEX0234_179 | | cyt | Myristyl 34-39; Pkc_Phospho_Site 30-32; | | |
| DEX0234_180 | | cyt | Pkc_Phospho_Site 19-21; | | |
| DEX0234_181 | | nuc | Ck2_Phospho_Site 33-36; Myristyl 46-51; Pkc_Phospho_Site 49-51;63-65;89-91; | | |
| DEX0234_182 | | nuc | | | |
| DEX0234_183 | 104-113 1.01 10 | pla | Ck2_Phospho_Site 87-90;141-144; Myristyl 115-120; Pkc_Phospho_Site 9-11;66-68;87-89;104-106; Tyr_Phospho_Site 89-96; | | 1 o31-53i |
| DEX0234_184 | 29-70 1.04 42 | exc | Asn_Glycosylation 62-65; Ck2_Phospho_Site 47-50; Myristyl 39-44;43-48;57-62; | 21 0.945 0.793 | |
| DEX0234_185 | 103-112 1.36 10 14-30 1.20 17 | nuc | Asn_Glycosylation 62-65;106-109; Camp_Phospho_Site 55-58;104-107; Ck2_Phospho_Site 41-44;50-53; Pkc_Phospho_Site 25-27;41-43;54-56;118-120; | | |
| DEX0234_186 | 53-65 1.23 13 | cyt | Amidation 699-702; Asn_Glycosylation 83-86;120-123;161-164;198-201;239-242;276-279;373-376;902-905; Atp_Gtp_A 527-534; Camp_Phospho_Site 701-704;876-879; Ck2_Phospho_Site 11-14;70-73;91-94;169-172;247-250;286-289;376-379;525-528;609-612;791-794;809-812;842-845; Myristyl 127-132;205-210;282-287;456-461;471-476;625-630;749-754;786-791;866-871; Pkc_Phospho_Site 84-86;130-132;162-164;208-210;240-242;414-416;540-542;681-683;754-756;770- | | |

| | | | | | |
|-------------|---|-----|---|--|-------------|
| | 744-755 1.19 12 864-880 1.11 17 564-581 1.06 18 588-604 1.06 17 435-460 1.03 26 697-709 1.02 13 271-281 1.02 11 | | 772;791-793;835-837;844-846;848-850; Transferrin_1 146-154;224-232; | | |
| DEX0234_187 | 40-62 1.14 23 | nuc | Myristyl 21-26; Pkc_Phospho_Site 82-84; | | |
| DEX0234_188 | | pla | Amidation 44-47; Pkc_Phospho_Site 9-11; | | 1 i25-43o |
| DEX0234_189 | 526-538 1.29 13 235-268 1.27 34 416-431 1.24 16 445-473 1.14 29 329-341 1.10 13 545-584 1.08 40 83-95 1.08 13 163-210 1.06 48 626-642 | pla | Amidation 99-102;129-132; Asn_Glycosylation 46-49;510-513;520-523;556-559; Ck2_Phospho_Site 83-86;122-125;214-217;242-245;263-266;333-336;357-360;376-379;380-383;392-395;439-442;455-458;459-462;465-468;487-490;498-501;512-515;516-519;543-546;552-555;558-561;604-607;611-614;628-631; Myristyl 50-55;57-62;90-95;179-184;260-265;264-269;267-272;324-329;372-377;472-477;548-553; Pkc_Phospho_Site 36-38;119-121;122-124;242-244;332-334;484-486;560-562;604-606; | | 1 i139-161o |

| | | | | | |
|-------------|---------|-----|--|--|--|
| | 1.05 17 | | | | |
| DEX0234_190 | | cyt | | | |

Example 6: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

5 RNA is isolated from individual patients or from a family of individuals that have a phenotype of interest. cDNA is then generated from these RNA samples using protocols known in the art. *See*, Sambrook (2001), *supra*. The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO: 1-100. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 10 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky *et al.*, *Science* 252(5006): 706-9 (1991). *See also* Sidransky *et al.*, *Science* 278(5340): 1054-9 (1997).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The 15 intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations are then cloned and sequenced to validate the results of the direct sequencing. PCR products is cloned into T-tailed vectors as described in Holton *et al.*, *Nucleic Acids Res.*, 19: 1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected 20 individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements may also be determined. Genomic clones are nick-translated with digoxigenin deoxyuridine 5' triphosphate (Boehringer Mannheim), and FISH is performed as described in Johnson *et al.*, *Methods Cell Biol.* 35: 73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 25 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C-and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and 30 variable excitation wavelength filters. *Id.* Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region

hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

Example 7: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample

5 Antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 ug/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described above. The wells are blocked so that non-specific binding of the polypeptide to the well is reduced. The coated wells are then
10 incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbound polypeptide. Next, 50 µl of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three
15 times with deionized or distilled water to remove unbound conjugate. 75 µl of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution are added to each well and incubated 1 hour at room temperature.

 The reaction is measured by a microtiter plate reader. A standard curve is prepared, using serial dilutions of a control sample, and polypeptide concentrations are plotted on
20 the X-axis (log scale) and fluorescence or absorbance on the Y-axis (linear scale). The concentration of the polypeptide in the sample is calculated using the standard curve.

Example 8: Formulating a Polypeptide

 The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the
25 individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

 As a general proposition, the total pharmaceutically effective amount of secreted
30 polypeptide administered parenterally per dose will be in the range of about 1 , µg/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most

preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 µg/kg/hour to about 50 mg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution
5 may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracisternally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an
10 oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

15 The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semipermeable polymer matrices in the form of shaped articles, e. g., films, or microcapsules. Sustained-release matrices include polylactides (U. S. Pat. No.3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22: 547-
20 556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15: 167-277 (1981), and R. Langer, Chem. Tech. 12: 98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE Epstein et
25 al., Proc. Natl. Acad. Sci. USA 82: 3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77: 4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U. S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected
30 proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable

carrier, i. e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation.

For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides. Generally, the
5 formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-
10 aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate,
15 succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e. g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates
20 including cellulose or its derivatives, glucose, manose, or dextrans; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will
25 be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e. g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container
30 having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampules or vials, as an aqueous solution or as a lyophilized formulation

for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1 % (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

5 The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container (s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or
10 sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 9: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or
15 normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the
20 activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 ug/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided above.

25

Example 10: Method of Treating Increased Levels of the Polypeptide

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

30 For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the

treatment was well tolerated. The formulation of the antisense polynucleotide is provided above.

Example 11: Method of Treatment Using Gene Therapy

5 One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down,
10 closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e. g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

 At this time, fresh media is added and subsequently changed every several days.
15 After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks. pMV-7 (Kirschmeier, P. T. et al., DNA, 7: 219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using
20 glass beads.

 The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and
25 the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB 101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

30 The amphotropic pA317 or GP+aml2 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now

produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing
5 the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media.

If the titer of virus is high, then virtually all fibroblasts will be infected and no
10 selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

15

Example 12: Method of Treatment Using Gene Therapy-In Vivo

Another aspect of the present invention is using in vivo gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences
20 into an animal to increase or decrease the expression of the polypeptide.

The polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the art, see, for example, W0 90/11092, W0 98/11779; U. S. Patent No. 5,693,622; 5,705,151; 5,580,859; Tabata H. et al. (1997) Cardiovasc. Res. 35 (3): 470-479, Chao J et al. (1997) Pharmacol. Res. 35 (6): 517-522, Wolff J. A. (1997) Neuromuscul. Disord. 7 (5): 314-318, Schwartz B. et al. (1996) Gene Ther. 3 (5): 405-411, and Tsurumi Y. et al. (1996) Circulation 94 (12): 3281-3290.

The polynucleotide constructs may be delivered by any method that delivers
30 injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, colon, liver, intestine and the like). The polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides of the present invention may also be
5 delivered in liposome formulations (such as those taught in Felgner P. L. et al. (1995) Ann. NY Acad. Sci. 772: 126-139 and Abdallah B. et al. (1995) Biol. Cell 85 (1): 1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain
10 sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the
15 desired polypeptide for periods of up to six months.

The polynucleotide construct can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, colon, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue.
20 Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to
25 the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin
30 fibroblasts. In vivo muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05 µg/kg body weight to about 50 mg/kg body

weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined
5 by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to colons or bronchial tissues, throat or mucous membranes of the nose. In addition, naked
10 polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle in vivo is determined as follows. Suitable template DNA for production of mRNA coding for polypeptide of the present invention is prepared in accordance with a standard
15 recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the
20 anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about 0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

25 After an appropriate incubation time (e. g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection
30 may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice.

The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA.

Example 13: Transgenic Animals

The polypeptides of the invention can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, hamsters, guinea pigs, pigs, micro-pigs, goats, sheep, cows and non-human primates, e. g., baboons, monkeys, and chimpanzees may be used to generate transgenic animals. In a specific embodiment, techniques described herein or otherwise known in the art, are used to express polypeptides of the invention in humans, as part of a gene therapy protocol.

Any technique known in the art may be used to introduce the transgene (I. e., polynucleotides of the invention) into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection (Paterson et al., Appl. Microbiol. Biotechnol. 40: 691-698 (1994); Carver et al., Biotechnology (NY) 11: 1263-1270 (1993); Wright et al., Biotechnology (NY) 9: 830-834 (1991); and Hoppe et al., U. S. Pat. No. 4,873,191 (1989)); retrovirus mediated gene transfer into germ lines (Van der Putten et al., Proc. Natl. Acad. Sci., USA 82: 6148-6152 (1985)), blastocysts or embryos; gene targeting in embryonic stem cells (Thompson et al., Cell 56: 313-321 (1989)); electroporation of cells or embryos (Lo, 1983, Mol Cell. Biol. 3: 1803-1814 (1983)); introduction of the polynucleotides of the invention using a gene gun (see, e. g., Ulmer et al., Science 259: 1745 (1993); introducing nucleic acid constructs into embryonic pluripotent stem cells and transferring the stem cells back into the blastocyst; and sperm mediated gene transfer (Lavitrano et al., Cell 57: 717-723 (1989). For a review of such techniques, see Gordon, "Transgenic Animals," Intl. Rev. Cytol. 115: 171-229 (1989).

Any technique known in the art may be used to produce transgenic clones containing polynucleotides of the invention, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (Campell et al., Nature 380: 64-66 (1996); Wilmut et al., Nature 385: 810813 (1997)).

The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, I. e., mosaic animals or chimeric. The transgene may be integrated as a single transgene or as multiple copies such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko et al. (Lasko et al., Proc. Natl. Acad. Sci. USA 89: 6232-6236 (1992)). The regulatory sequences required for

such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the polynucleotide transgene be integrated into the chromosomal site of the endogenous gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some
5 nucleotide sequences homologous to the endogenous gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type, by following, for example, the teaching of Gu et
10 al. (Gu et al., Science 265: 103-106 (1994)). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished
15 by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (rt-PCR).
20 Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding
25 strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to
30 both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Example 14: Knock-Out Animals

Endogenous gene expression can also be reduced by inactivating or "knocking out" the gene and/or its promoter using targeted homologous recombination. (E. g., see Smithies et al., Nature 317: 230-234 (1985); Thomas & Capecchi, Cell 51: 503-512 (1987); Thompson et al., Cell 5: 313-321 (1989)). For example, a mutant, non-functional polynucleotide of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous polynucleotide sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention in vivo. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene (e. g., see Thomas & Capecchi 1987 and Thompson 1989, supra). However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site in vivo using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (e. g., knockouts) are administered to a patient in vivo. Such cells may be obtained from the patient (I. e., animal, including human) or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (e. g., lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered in vitro using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence

associated with the polypeptides of the invention, e. g., by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

5 The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, e. g., in the circulation, or intraperitoneally.

10 Alternatively, the cells can be incorporated into a matrix and implanted in the body, e. g., genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. (See, for example, Anderson et al. U. S. Patent No. 5,399,349; and Mulligan & Wilson, U. S. Patent No. 5,460,959 each of which is incorporated by reference
15 herein in its entirety).

 When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of
20 components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

 Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with
25 aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

 While preferred illustrative embodiments of the present invention are described, one skilled in the art will appreciate that the present invention can be practiced by other than the described embodiments, which are presented for purposes of illustration only and
30 not by way of limitation. The present invention is limited only by the claims that follow.

What is Claimed is:

1. An isolated nucleic acid molecule comprising:
 - (a) a nucleic acid molecule comprising a nucleic acid sequence that encodes an amino acid sequence of SEQ ID NO: 109, 119, 120, 122, 123, 124, 125, 153, 166, 167, 182, 183, 184, 185, 189 or 190;
 - (b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 10, 20, 21, 24, 25, 26, 27, 58, 73, 74, 92, 93, 94, 95, 99 or 100;
 - (c) a nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of (a) or (b); or
 - (d) a nucleic acid molecule having at least 90% sequence identity to the nucleic acid molecule of (a) or (b).
2. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is a cDNA.
3. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is genomic DNA.
4. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is a mammalian nucleic acid molecule.
5. The nucleic acid molecule according to claim 4, wherein the nucleic acid molecule is a human nucleic acid molecule.
6. A method for determining the presence of a colon specific nucleic acid (CSNA) in a sample, comprising the steps of:
 - (a) contacting the sample with the nucleic acid molecule of SEQ ID NO: 9, 10, 11, 20, 21, 24, 25, 26, 27, 30, 31, 58, 72, 73, 74, 92, 93, 94, 95, 96, 99 or 100 under conditions in which the nucleic acid molecule will selectively hybridize to a colon specific nucleic acid; and
 - (b) detecting hybridization of the nucleic acid molecule to a CSNA in the sample, wherein the detection of the hybridization indicates the presence of a CSNA in the sample.

7. A vector comprising the nucleic acid molecule of claim 1.

8. A host cell comprising the vector according to claim 7.

5

9. A method for producing a polypeptide encoded by the nucleic acid molecule according to claim 1, comprising the steps of:

- (a) providing a host cell comprising the nucleic acid molecule operably linked to one or more expression control sequences, and
- 10 (b) incubating the host cell under conditions in which the polypeptide is produced.

10. A polypeptide encoded by the nucleic acid molecule according to claim 1.

15 11. An isolated polypeptide selected from the group consisting of:

- (a) a polypeptide comprising an amino acid sequence with at least 60% sequence identity to of SEQ ID NO: 109, 119, 120, 122, 123, 124, 125, 153, 166, 167, 182, 183, 184, 185, 189 or 190 ; or
- (b) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule having at least 90% sequence identity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 10, 20, 21, 24, 25, 26, 27, 58, 73, 74, 92, 93, 94, 95, 99 or 100.

25 12. An antibody or fragment thereof that specifically binds to

- (a) a polypeptide comprising an amino acid sequence with at least 60% sequence identity to of SEQ ID NO: 108, 109, 110, 119, 120, 122, 123, 124, 125, 127, 128, 153, 165, 166, 167, 182, 183, 184, 185, 186, 189 or 190; or
- (b) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule having at least 90% sequence identity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 9, 10, 11, 20, 21, 24, 25, 26, 27, 30, 31, 58, 72, 73, 74, 92, 93, 94, 95, 96, 99 or 100.

13. A method for determining the presence of a colon specific protein in a sample, comprising the steps of:

- (a) contacting the sample with a suitable reagent under conditions in which the reagent will selectively interact with the colon specific protein comprising an amino acid sequence with at least 60% sequence identity to of SEQ ID NO: 108, 109, 110, 119, 120, 122, 123, 124, 125, 127, 128, 153, 165, 166, 167, 182, 183, 184, 185, 186, 189 or 190; and
- (b) detecting the interaction of the reagent with a colon specific protein in the sample, wherein the detection of binding indicates the presence of a colon specific protein in the sample.

10

14. A method for diagnosing or monitoring the presence and metastases of colon cancer in a patient, comprising the steps of:

- (a) determining an amount of the nucleic acid molecule of claim 1 or a polypeptide of claim 11 in a sample of a patient; and
- (b) comparing the amount of the determined nucleic acid molecule or the polypeptide in the sample of the patient to the amount of the colon specific marker in a normal control; wherein a difference in the amount of the nucleic acid molecule or the polypeptide in the sample compared to the amount of the nucleic acid molecule or the polypeptide in the normal control is associated with the presence of colon cancer.

20

15. A kit for detecting a risk of cancer or presence of cancer in a patient, said kit comprising a means for determining the presence the nucleic acid molecule of claim 1 or a polypeptide of claim 11 in a sample of a patient.

25

16. A method of treating a patient with colon cancer, comprising the step of administering a composition according to claim 11 or 12 to a patient in need thereof, wherein said administration induces an immune response against the colon cancer cell expressing the nucleic acid molecule or polypeptide.

30

17. A vaccine comprising the polypeptide or the nucleic acid encoding the polypeptide of claim 11.

1

SEQUENCE LISTING

<110> diaDexus, Inc.
Sun, Yongming
Liu, Chenghua
Ghosh, Malavika

<120> Compositions and Methods Relating to Colon Specific Genes and
Proteins

<130> DEX-0342

<150> US 60/316,259

<151> 2001-08-31

<160> 190

<170> PatentIn version 3.1

<210> 1

<211> 461

<212> DNA

<213> Homo sapien

<220>

<221> misc_feature

<222> (288)..(288)

<223> n=a, c, g or t

<220>

<221> misc_feature

<222> (290)..(290)

<223> n=a, c, g or t

<220>

<221> misc_feature

<222> (393)..(393)

<223> n=a, c, g or t

<400> 1

agcaatatac taagggttac cttgccctca gtgggccag agtagatgac acgatgtggg 60

ataagaagtg gcccatgcac agaaggagag actccttttg tctggagaca ctatggcgag 120

tttcttagac aagggtggtat ccgagatgag tttgcaaacg tgtgcacgat ttgatcatgc 180

agatgtggca ggaagagcat tccagatagg aggaatcacg tgaggatgag tatggtggca 240

ggaaagggag aaccatatat atataagaga ggaaaggaag tgttaganan ggtgttgcac 300

gcatgaaagg gaagaggagg aagaaatgaa gttggcaaga caaatctgaa cctgatctgg 360

gagaatcttg actatctcat gatttttgag agntgggtgca caaacatct gtcttggtta 420

cttgctggaa tgttgtggcc tgtctggttt gctttttaga c 461

<210> 2

2

<211> 943
 <212> DNA
 <213> Homo sapien

<400> 2
 ggatgggccc cagttccagc cttggctttg cgtcctgacc agtgtgtctg tgggcaagtc 60
 acctgacatt tcttgtcttc aagtctcatt tgtgaagtga agaggggtgac gaaccacaca 120
 gataacaaga ggttcttttt agtgtaaga gttatgtaac taatagccat ccctctttga 180
 ttcttcttgc actacttgca gggtagcagt gagaaaacta agaagaaatt ttgagttaga 240
 agttgagtga attcttggtc ctttcatggc taaaatatgt gaccaccagc ccactgaatg 300
 aaatggagcg tgtttgaatt tgcctcactg tggtcactct ggaagcaggc tcagcttacc 360
 agggccattt ctctgatgta cctttggaat tcttccaaaa tgtataggga gaattagaaa 420
 aaccagaggc attctgagtg aggaaataag cagatgactc agcaaatac caatctactt 480
 catctccttg aatagcaaat agttgaattt catgaagaaa ctttaatttt ttctgattat 540
 gaaagtcata tacaattggt ttaaccaagt gtgcacagaa tgatcacact tatagaaaat 600
 gtctgttttg ttcataagtg aattaaag gtcaggaaag atatattcca tattatgaac 660
 atttttttta gaaaatgaaa tgggcatgat gattagaaaa atgttgtaag tgctcattgt 720
 aggacagtca gataatatat aaagatgcaa aaaataaaaa ataaaaatca ccataatcc 780
 tatgccctaa taacaatac agtttgattt tggcataatt tccctcaagt cagttttcct 840
 ctgcttagaa aaaggggttg ggaaaataga tactgtctac tcaattgtgt gtcttttatt 900
 tttaaactta tgtcttaagc actttacact ggattagaat ccc 943

<210> 3
 <211> 537
 <212> DNA
 <213> Homo sapien

<400> 3
 gctcgagggtg aacgtcatta aattaaagc tagaaattat taagctaagt gagaaaggca 60
 tgccaagtaa gtataaatcc ggagttgtcc aacttgagct gaagtgctat tagatctacc 120
 aagtcataag actgaatggt cactgcaacc atctatccta taatggatgc aatatgtttt 180
 agggcaagag aaaatgcaaa agacaaaagt aggctgaatg aacagttggc ccagatccct 240
 atatcactga tgccctcttc tcagcttata catacagtct cctaggtcct tccaaccagc 300
 tgatacctta gtaaacaaagc caggcctaata gcaagggcac tttggctagg tatattggca 360
 caatatgtca accactcaga tgtgggacct caaatataat ccatgagggt aaacctactc 420
 tgataagtag aatacataga ttcacgcca tccatagaaa gggttcagaa aacaggagct 480
 ggccctgtgc ttttacagag cttttatcctt ccataaggg caggaccaga tattatc 537

<210> 4
 <211> 513
 <212> DNA
 <213> Homo sapien

 <220>
 <221> misc_feature
 <222> (129)..(319)
 <223> n=a, c, g or t

<400> 4
 ggggtagtaa atgcatagta aaatacatag ttgtttgctt ttaaccacgc tgggagtgca 60
 gtaggattgt ttacagcagc aatcccacaa acatgtaagt aatgcattgt gctacaatat 120
 tattacggnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 180
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240
 nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300
 nnnnnnnnnn nnnnnnnnnn ccttatgcaa aacagtgcga gaagttatac actgcaataa 360
 gacaagaaaa taattaaagg acatacagat cagaatggaa aaaatagaac agtccttacc 420
 tacaacagac atgactttct acatagaaat ccaaataaat ctatttcaca aacctcctta 480
 aattaagaag tgagtccatt aaggtcacag aat 513

<210> 5
 <211> 255
 <212> DNA
 <213> Homo sapien

<400> 5
 ttaatcaaat ttcaaaacca gggtttgtag tacatttaaa ttgcatattc caaagcagtt 60
 ggggtttgcct gcgttcagct ttaatattaa gctatacttc cctttcaaat aagggtatatt 120
 catcgtaaag cctgtaaatt ctagtgtgtc attgtttaga tatttatagt cattttaata 180
 tatctgttta cgccagctg caatggctaa cacctgtaaa ctcagcactt ttttgaggcc 240
 aagggtggggc cgatt 255

<210> 6
 <211> 777
 <212> DNA
 <213> Homo sapien

<400> 6
 gctgttaaca tcaagaggtg gggatcatag gaggtccatc ataaagtttg tcctccaccg 60
 tagaggggag gtgacgggat tagaaggcat tataacagtt cttcaggtga ggccttgagg 120
 gatgcagatg gaaaaacggg agctttcaga ggcggaagtt ttggaagtgc agttccagtt 180

gaaagttggg atagtcaggt ggtgacgctt ctcaggaggg cttctgtgcc ctttgtcttc 240
 atgtgccttc tatgtgctct ctctgttag acaaatgaac attctgtgtt acagttgcca 300
 ccgataaatc catctgtctg gctaaaatgg aattatttac cagccaattc ctggccagtg 360
 gaggtatctg ccctcttgcc ttatggtttt gttgaggatc aaatgaggta attcatatga 420
 agtgttttat agactacaaa gtaatatgca aacaaaaggt attaaacggg gctacctttt 480
 tctctatgtt aagtagaacc catggaattt aacttacatt ttccttaata acttgggttc 540
 attatgttgt atgtagctta ttaaactggg ctttatTTaa cactgtgatt tgggtttttg 600
 gagtagagag ttacttttta tgataaatga ccccaaactg ctataccttt caaatccttg 660
 tgtcccatTT ttattgttta aaaaaacaat ctcaccttgc cgttgcctgt cactcttttg 720
 cttctgtgtg tcagttttga gctacatttt tccccctcg atttggtgct tctgac 777

<210> 7
 <211> 234
 <212> DNA
 <213> Homo sapien

<400> 7
 tgtttgtaaa ttgtttgtgt aattaaaaat taacttttaa aaacaaaatt ggaaaactat 60
 agtacgatcc aattttggct aaaaatatag ctacagcaat atatataccg agaaaagcgg 120
 ggaatacata ttgaattgtt gaaagtgtgt atctctgcat agggtagtga caggatggga 180
 ttatgggagg attttgactt cctgtgacgg gtgttagggc ctgaatatTT ctgt 234

<210> 8
 <211> 1484
 <212> DNA
 <213> Homo sapien

<400> 8
 acagtattct tggacaatta atgggaagtt tcagcaatca ggacaaaatc tgtttatccc 60
 ccaaattact acaaagcata gcgggctcta tgtttgctct gttcgtaact cagccactgg 120
 cgaggaaagc tccacatcgt tgacagtcaa agtctctgct tctacaagaa taggacttct 180
 tcctctcctt aatccaacat agcagctgtg atgtcatttc tgtatttcag gaagactggc 240
 agagacgtgg tctccctttc ttgctcaggc tggctctaaa ctcttggttt caagtgatcc 300
 tcctgcctta gagtctcaaa ctgctgggat tacaggcctc tgccagcaca cctggccccc 360
 caattcttca aaagctttca gagaaagcat tgtagaagac atgactttgt gactgtaagt 420
 ttcactgat octacatcac taggaaggct cattcttaag aagtcatgca tggaaaaggg 480
 gatgaagaca aatcagaaaa gggcaaaggg aagtcacagc aaaaaaggga cagtataatc 540

5

```

ctggaacctt attaaagtca cacaactgct gcttcaatta gagcaattcg tttggcaaac 600
atcgttctaa ccctatagac taggttttcc agagtttctg aagtatcttc taattgcaat 660
ggcaatctga cacatthttc tgaattgcag tctgaatcca gtgttcattg gtaccttttg 720
tatagtccac acatcagcag ccacaaagg tgtttatata taagttgcta tgatttctcc 780
agaagtttac ataagtcatt gagtttcagc ttgcaagggt taatctatcc atctgacaaa 840
aggctaatat ccaaaatcta caagaaactt aaacaaattht acaagaaaaa aaaaaaaaaa 900
aaaatttggg ggggggttggg cccggaaaag gtttttaaaa ctctgggtggg gcgcgcgggg 960
gcgcaaaaag aagggttacg gtgaaccact gtgggccatt agcgtgtgtt ccctaaggag 1020
aagactccgg gcgcaagaga gccaaagggg cgtggggtct ctctcccaa tataaaaaac 1080
caggccgggg ggggtaccga gcgcgggttc tgagaacaac taccaaaagg ggggcttctc 1140
agggggcgac actaaggggg tctctatctc acaagagggg gctctcaaga ggggagccgt 1200
ggaatatctt caataaattht aagccccgg ctctgtgtgg gcttactccc actatcacta 1260
gccagtccga taaaacaaca gcaccgcgag caaacatcac actaagtact aggcacgcga 1320
tccacgacca cgcaatcaa taccaccgag gatctgtcaa ccaagcgagg agcagtacag 1380
agcacgacca aactgctagc accacaaccg catccagaag caaacgtaca gcggcaaaga 1440
gcagagcacc accaacagga ggtgcaaaga gagaagcaga ccaa 1484

```

```

<210> 9
<211> 5046
<212> DNA
<213> Homo sapien

```

```

<400> 9
atggcagaag gcaaggagaa gcaagtcaca tcttacatgg atggcagcag accttatgac 60
gtgagcatga cctacatcca caaagcagga gggcctgac agcaggaatt agtcatgtct 120
acctgcactg tgccccttga ctctgtgtg catctgcctc agggccgaac gaattatcgc 180
aaatatttcc ggtcagaggc tgccttcacc ttggcagatt tcatctacaa gagcatgatc 240
cgagtaaata gttccagatt ggttcgagtc acccaggtgg agaattgagga gaaactgaag 300
gagctagagc agtttagtat ctggaacttht ttttctctct ttttaaaaga gaaattgaat 360
gacacctatg ttaacgtggg tctatacagc acaaaaacct gcctcaaagt tgagattata 420
gagaaggaca ccaagtacag tgtcattgtg atccggagggt cctgggatgt tatccgagta 480
aatagttcca gattgggttc agtcaccag gtggagaatg aggagaaact gaaggagcta 540
gagcagttta gtatctggaa ctttttttcc tcctttttta aagagaaatt gaatgacacc 600
tatgttaacg tgggtctata cagcacaana acctgcctca aagttgagat tatagagaag 660

```


| | |
|--|------|
| gacaccaagt acagtgtcat tgtgatccgg aggtcctggg atgttatccg agtaaatagt | 720 |
| tccagattgg ttcgagtcac ccaggtggag aatgaggaga aactgaagga gctagagcag | 780 |
| tttagtatct ggaacttttt ttctctcttt ttaaaagaga aattgaatga cacctatgtt | 840 |
| aacggcatac cgtggacaaa ggtggactac tttgataatg gcatcatttg taagctcatt | 900 |
| gagcataatc agcgaggat cctggccatg ttggatgagg agtgccctgcg gcctggggtg | 960 |
| gtcagtgact ccactttcct agcaaagctg aaccagctct tctccaagca tggccactac | 1020 |
| gagagcaaag tcacccagaa tgcccagcgt cagtatgacc acaccatggg cctcagctgc | 1080 |
| ttccgcactc gccactatgc gggcaagggtg acatacaacg tgaccagctt tattgacaag | 1140 |
| aataatgacc tactcttccg agacctgttg caggccatgt ggaaggccca gcacccctc | 1200 |
| cttcggtcct tgtttcctga gggcaatcct aagcaggcat ctctcaaacg cccccgact | 1260 |
| gctggggccc agttcaagag ttctgtggcc atcctcatga agaactctga ttccaagagc | 1320 |
| cccaactaca tcaggtgcat aaagcccaat gagcatcagc agcgaggta gttctcttca | 1380 |
| gacctggtgg caaccaggc tcggtacctg ggactgctgg agaacgtacg ggtgcgacgg | 1440 |
| gcaggctatg cccaccgcca gggttatggg cccttctctg aaaggtagcg attgctgagc | 1500 |
| cggagcacct ggccctactg gaatggggga gaccgggaag gtgttgagaa ggtcctgggg | 1560 |
| gagctgagca tgtcctcggg ggagctggcc tttggcaaga caaagatctt cattagaagc | 1620 |
| cccaagactc ttttctacct cgaagaacag aggcgcctga gactccagca gctggccaca | 1680 |
| ctcatacaga agatttaccg aggctggcgc tgccgcaccc actaccaact gatgcgaaag | 1740 |
| agtcagatcc tcatctctc ttggtttcgg ggaacatgg cccgaaagaa ttatcgcaaa | 1800 |
| tatttccggt cagaggctgc cctcaccttg gcagatttca tctacaagag catggtacag | 1860 |
| aaattcctac tggggctgaa gaacaatttg ccatccacaa acgtcttaga caagacatgg | 1920 |
| ccagccgccc cctacaagt cctcagcaca gcaaactcagg agctgcagca gctcttctac | 1980 |
| cagtgggaag caactcctgt ccctccttcc tcacagtga agaggttccg ggatcagctg | 2040 |
| tccccgaagc aggtagagat cctgagggaa aagctctgtg ccagtgaact gttcaagggc | 2100 |
| aagaaggctt catatcccca gagtgtcccc attccattct gtggtgacta cattgggctg | 2160 |
| caagggaacc ccaagctgca gaagctgaaa ggcggggagg aggggcctgt tctgatggca | 2220 |
| gaggccgtga agaaggtaaa tcgtggcaat ggcaagactt cttctcggat tctcctcctg | 2280 |
| accaagggcc atgtgattct cacagacacc aagaagtccc aggccaaaat tgtcattggg | 2340 |
| ctagacaatg tggctgggggt gtcagtcacc agcctcaagg atgggctctt tagcttgcac | 2400 |
| ctgagtgaga tgtcatcggt gggctccaag ggggacttcc tgctggtcag cgagcatgtg | 2460 |
| attgaactgc tgaccaaaat gtaccgggct gtgctggatg ccacgcagag gcagcttaca | 2520 |

| | | | | | | |
|------------|-------------|------------|------------|------------|------------|------|
| gtcacctgga | ctgagaagtt | ctcagtgagg | ttcaaggaga | acagtgtggc | tgtcaaggtc | 2580 |
| gtccagggcc | ctgcaggtgg | tgacaacagc | aagctacgct | acaaaaaaaa | ggggagtcac | 2640 |
| tgcttgagg | tgactgtgca | gcagctcacg | gccggctatc | atgctggcca | gggagaactg | 2700 |
| attaacttct | cttcctgcct | gcagattaat | ctgctgtctg | agcacaagcc | acgtgcttct | 2760 |
| ggcacaccct | gctttgagct | gagatagaac | ctggggaatc | atctgttttc | aggcgggtga | 2820 |
| ggggctagag | cctgccttgt | ttgggaggag | ggtggctctg | ttcagaatag | gggtagctca | 2880 |
| ggctctggcc | agccttctcc | cgcctccaac | agctccccc | atccttgact | tctcagaatc | 2940 |
| aggccgagaa | gagcctatct | ggccgagagt | gggggtgtga | cctgcgcctc | atcgcccccg | 3000 |
| ctctccatct | catctcctgc | tcccagggcc | caaattgtcg | tcactttccc | agtgaagtgt | 3060 |
| ctggtcattt | tcagaagcaa | tttcaggaga | acatgcagct | gccgctccct | atcctgcatt | 3120 |
| tcccttcaca | gggctgaagg | cactgtcagc | tccctgggct | gggggtgatg | ggagagggga | 3180 |
| agggctaggg | ccctcaccct | tgctctcact | gtgcccatca | tgtagatgga | ctggagttca | 3240 |
| aggaagggca | ggcactcccc | tcctccttta | ctcttctgtc | actctcttcc | tcctcttctt | 3300 |
| tcctgtctct | gcctctcttt | tctggagcct | aggagtgtgt | gttttcatcc | cctgaaacaa | 3360 |
| atagggactc | agtttcccca | cctgtgttac | aggggtggaa | ttggctccat | cactgtggga | 3420 |
| gaagctggag | ttctgtctacc | agtcctcccc | tccccagccc | tgctcttctt | ctcccagccc | 3480 |
| tctcccttca | gccagttcag | cgtcttgaga | gtctgggttg | tttcagcctc | tgaggggcac | 3540 |
| aagccatcct | ggattccctt | aaccccatga | ggagccattc | tagcatctca | cagcttaaac | 3600 |
| cagctctagc | tcagtctctc | tggtctagtc | catttttctt | cctcaggctc | tgagggcctc | 3660 |
| ttgttccttg | ctctgtgggg | ttttctccag | ttgtctcctg | gctgcaggac | atggcaggac | 3720 |
| atagaatgct | gtcatccttc | cactcttcat | tggcatctcc | accagtgctc | acatatgacc | 3780 |
| ctagccctgg | ctctccctct | gccagtacct | cctctgggat | tttgcgagag | tccacaagtt | 3840 |
| gtgcatgtgg | tggatatatt | caggccatct | tgtgtgtaca | agcctagagg | gtctgcttcc | 3900 |
| acctctggcc | ctcagtgaat | tgtgtactaa | cctgtctcaa | cacagcacia | ctgtacacac | 3960 |
| cttttcctgg | cctcatccct | aacccatcat | agcagcaaag | aggggaagtt | gcaggggagg | 4020 |
| agctgctaag | gaccctggac | tccaagtacc | ctgctcctct | aggccaggga | catcatctga | 4080 |
| gatgtggctc | aaataaagg | tgggtgttca | agaaaaaaca | cttggggact | ctatagctgc | 4140 |
| aacaccact | ttacatgtca | tttccatatg | atttgtaggc | aaaatgaagc | ccaggctgtc | 4200 |
| ctagccctcc | aatacctccc | tctctcatca | cctctccaac | atagcctagc | attagctctt | 4260 |
| tcaagtcttt | gctaatacca | gagatcaagg | ggtgatcaac | tctccctgcc | atccccttgt | 4320 |

```

tccccgcacc ccccgccccg gctccccccac catccttggc tcctgccatc ctctttgaga 4380
tgctgcatca tcaaaggaca ttatttatgg tgtacctttg ctgaagccct gcttccttgg 4440
tgccagggct tgggagcagg gatgggtggg ttgggtgggg agaggggtgg atgcagagat 4500
tggacccagg aggcttttag tcctcagctc ttggcttaac acctcctcct cttacacacc 4560
caactccctc cagcctgccc agcttgggcc ttcagctcca gattgggtgg gttaggagag 4620
gaggaggagg gagatggatg gaaccaatta ggaacagcac ctgggctcct cacaggaatg 4680
aaccagtcat gccatttgca tgtaaacagc ttcccacttc tctcctcatc ctaccaaagt 4740
ctcccaaccc tgggttcttg cccatgttct ttgccacac agccctgtaa ttagctgggt 4800
aatgagaagc ttttaatgag tcccattagc atctcgtgta ataaagaggc cttgagaccc 4860
agctgctgtc ctcacttttg gatgaacacg ggtccctgtg tagccagtga cttctgtcag 4920
tacagtctaa gttctcggat ggggtgggag acaaacattt caggacccca gcagcacttg 4980
agaggttcca tgggtgatcc atgtttttga ctgtgataca agaaacttgg ctctggcttc 5040
cttggtt 5046

```

```

<210> 10
<211> 1493
<212> DNA
<213> Homo sapien

```

```

<400> 10
acacacgccc cagtctcca gtcctccgtg cgcgtcgggc ccgagccac agagtccatg 60
gaacccacgg agccgatgga acctacggag cccatggaac ctacggagcc catggaacct 120
acggagccca tggaaacctac ggagcccatg gaaccggcgc ggagcgcgca ccgtgggggc 180
gaggcgctgc tgcgggagct ggaggtgctg gtgcaggacg tggtagggac gagctcctgg 240
tgggagcgcc acggcggtgga ctgcgccatc ctgcgctca gcctcttcgc cttgccggca 300
ggcttcctgt gcctgcgctg ggagaatgcc ctggctcttg catccggcat caccatcttg 360
ggtgtgtgcc actacacact cactgtcaag ggcagccacc tggccactca tggggccctc 420
accgagtcca aacgctggag caagatctgg ctgcttttct ttgtggaggt gtgcacagcc 480
ttcactgcag agcacgccac gcatgggcac gtcaagatgc accatgccta caccaacgtg 540
gtgggccttg gggactccag cacgtggagg ctgccttgcc tcaaccgcta tgtctacatg 600
ttccttgctc ctttcctcct ccccatcgcc actccactgg tggctgtcga gcggctgagg 660
aagtgaggac tcgggacagc cctgcggacg ctggccctga tttctctggg cctttattct 720
cactactggc tgctcctgaa cgtgtcaggc ttcaagaacc ccagctcagc cctgggctgc 780
atgttctca ccagatccct gttggcccac ccctacctcc acgtcaacat cttccagcac 840

```

atcggactgc ccatgttctc ccgggacaac aagccccgtc ggattcacat gatgagcctg 900
 ggggtgctta acctggcccg gctgcccgtg ctggactggg cgttcggcca ctcgatcatc 960
 agctgccatg tggaacacca tctattcccc aggcctctctg ataacatgtg cctgaaggta 1020
 gtcgagggct gggctggggg cgctggaata aagggactgt tagaggatgg aaaggaagat 1080
 tcttacggcc tgggggcgtt gctcacacta taatcctggc acttgccggag gctgaggtgg 1140
 gcagatcgct ttgagcccag gagtatgaga gcagctgggc aacatggtga aaccccatct 1200
 ctacaaaaaa ttcaaaaatt agcgggggtgt ggtggcacat gcttgtggtc caagctgttg 1260
 gaagctgaag tgaggaccaa ctgagctctgc agaggttgag gctgcatgga accataatcg 1320
 aggcactgca ctacagccgg gtgacagagt aggaccctgt ctcaacaaaa gaacacaaag 1380
 gggcgaccgc caggaaacac cctgcagggg ccaagataag ggaccagaat ttagacaag 1440
 ggcgccaatg ggagagcaaa agagagaaaag gcggagaaaa acaggggacg gaa 1493

<210> 11
 <211> 3892
 <212> DNA
 <213> Homo sapien

<400> 11
 tcgagcctgg cggtagaatc ttcccagtag gcggcgccgg agggaaaaga ggattgaggg 60
 gctaggcccg gcggtatccc tctcccccg atgtgagcag ttttccgaaa ccccgtcagg 120
 cgaaggctgc ccagagaggt ggagtcggta gcggggccgg gaacatgagg cagtctctcc 180
 tattctctgac cagcgtgggt cctttctgtc tggcgccgg acctccggat gaccggggct 240
 tcggccccca ccagagactc gagaagcttg attctttgt ctcagactac gatattctct 300
 ctttatctaa tatccagcag cattcggtaa gaaaaagaga tctacagact tcaacacatg 360
 tagaaacact actaactttt tcagctttga aaaggcattt taaattatac ctgacatcaa 420
 gtactgaacg tttttcacia aatttcaagg tcgtgggtgt ggatggtaaa aacgaaagcg 480
 agtacactgt aaaatggcag gacttcttca ctggacacgt ggttggtgag cctgactcta 540
 gggttctagc ccacataaga gatgatgatg ttataatcag aatcaacaca gatggggccg 600
 aatataacat agagccactt tggagatttg ttaatgatac caaagacaaa agaattgttag 660
 tttataaatc tgaagatata aagaatgttt cacgtttgca gtctccaaaa gtgtgtgggt 720
 atttaaaagt ggataatgaa gagttgctcc caaaaggggt agtagacaga gaaccacctg 780
 aagagcttgt tcatcgagtg aaaagaagag ctgaccacga tcccatgaag aacacgtgta 840
 aattattggt ggtagcagat catcgcttct acagatacat gggcagaggg gaagagagta 900
 caactacaaa ttacttaata gagctaattg acagagttga tgacatctat cggaacactt 960

| | |
|--|------|
| catgggataa tgcagggtttt aaaggctatg gaatacagat agagcagatt cgcattctca | 1020 |
| agtctccaca agaggtaaaa cctgggtgaaa agcactacaa catggcaaaa agttacccaa | 1080 |
| atgaagaaaa ggatgcttgg gatgtgaaga tgttgctaga gcaatttagc tttgatatag | 1140 |
| ctgaggaagc atctaaagtt tgcttggcac accttttcac ataccaagat tttgatatgg | 1200 |
| gaactcttgg attagcttat gttggctctc ccagagcaaa cagccatgga ggtgtttgtc | 1260 |
| caaaggctta ttatagccca gttgggaaga aaaatatcta tttgaatagt ggtttgacga | 1320 |
| gcacaaagaa ttatggtaaa accatcctta caaagggaagc tgacctggtt acaactcatg | 1380 |
| aattgggaca taattttggg gcagaacatg atccggatgg tctagcagaa tgtgccccga | 1440 |
| atgaggacca gggagggaata tatgtcatgt atcccatagc tgtgagtggc gatcacgaga | 1500 |
| acaataagat gttttcaaac tgcagtaaac aatcaatcta taagaccatt gaaagtaagg | 1560 |
| cccaggagtg ttttcaagaa cgcagcaata aagtttgtgg gaactcgagg gtggatgaag | 1620 |
| gagaagagtg tgatcctggc atcatgtatc tgaacaacga cacctgctgc aacagcgact | 1680 |
| gcacgttgaa ggaagggtgc cagtgcagtg acaggaacag tccttgctgt aaaaactgtc | 1740 |
| agtttgagac tgcccagaag aagtgccagg aggcgattaa tgctacttgc aaaggcgtgt | 1800 |
| cctactgcac aggtaatagc agtgagtgcc cgctccagg aaatgctgaa gatgacactg | 1860 |
| tttgcttggg tcttggcaag tgtaaggatg ggaaatgcat ccctttctgc gagagggaac | 1920 |
| agcagctgga gtctgtgca tgtaatgaaa ctgacaactc ctgcaagggtg tgctgcaggg | 1980 |
| acctttctgg ccgctgtgtg ccctatgtcg atgctgaaca aaagaactta tttttgagga | 2040 |
| aaggaaagcc ctgtacagta ggattttgtg acatgaatgg caaatgtgag aaacgagtac | 2100 |
| aggatgtaat tgaacgattt tgggatttca ttgaccagct gagcatcaat acttttggaa | 2160 |
| agtttttagc agacaacatc gttgggtctg tcctgggttt ctccttgata ttttggaattc | 2220 |
| ctttcagcat tcttgtccat tgtgtgataa gaaattggat aaacagtatg aatctctgtc | 2280 |
| tctgtttcac cccagtaacg tcgaaatgct gagcagcatg gattctgcat cggttcgc | 2340 |
| tatcaaacc tttctgctgc cccagactcc aggcgcctg cagcctgccc ctgtgatcct | 2400 |
| tcggcgccag cagctccaaa actggaccac cagagaatgg acaccatcca ggaagacccc | 2460 |
| agcacagact cacatatgga cgaggatggg tttgagaagg accccttccc agcaatagcg | 2520 |
| acagctgcca agtcatttga ggatctcacg gaccatccgg tcaccagaag tgaaaaggct | 2580 |
| gcctccttta aactgcagcg tcagaatcgt gttgacagca aagaaacaga gtgctaattt | 2640 |
| agttctcagc tcttctgact taagtgtgca aaatatTTTT atagatttga cctacaaatc | 2700 |
| aatcacagct tgtattttgt gaagactggg aagtgactta gcagatgctg gtcattgtgt | 2760 |
| tgaacttcct gcaggtaaac agttcttgtg tggtttggcc cttctccttt tgaaaaggta | 2820 |

```

agggtgaaggt gaatctagct tatTTTtgagg cTTtcaggtt ttagtTTTta aaatatcttt 2880
tgacctgtgg tgcaaaagca gaaaatacag ctggattggg ttatgaatat ttacgtTTTT 2940
gtaaattaat cTTTtatatt gataacagca ctgactaggg aaatgatcag TTTTttTTta 3000
tacctgtaa tgaaccgctg aatatgaggc atttggcatt tatttgtgat gacaactgga 3060
atagtTTTTt TTTTttTTTT tTTTttTgcc ttcaactaaa aacaaaggag ataaatctag 3120
tatacattgt ctctaaattg tgggtctatt tctagtatt acccagagtt tttatgtagc 3180
agggaaaata tatatctaaa tttagaaatc atttgggtta atatggctct tcataattct 3240
aagactaatg ctctctagaa acctaaccac ctaccttaca gtgagggcta tacatggtag 3300
ccagttgaat ttatggaatc taccaactgt ttagggccct gatttgcctg gcagTTTTtc 3360
tgtatTTTat aagtatcttc atgtatccct gttactgata gggatacatg ctcttagaaa 3420
attcactatt ggctgggagt ggtggctcat gcctgtaatc ccagcacttg gagaggctga 3480
ggttgcgcca ctacactcca gcctgggtga cagagtgaga tcctgtctca aaaaaaaaaa 3540
aaatatcacc cttaaataca caaaaagacc tcatgtcat gcagtggaca tttcttggcc 3600
tgTTTTctg tgggtTTTTt tcagcatgat atgaagcaac tgttggcata gccagtggga 3660
taaataagaa atggaatcat ttggctgcct cgtcacttg gctaattggtc agttgtacac 3720
acatgacatg tggaagtga cgttagcagc ggtcactcat tttacaatt gaactctgct 3780
gggttgggta tttaccttg tctagaaaag aaatatgata tttttctct aaataaattt 3840
atttcatca ttcaaaaaa aaaaaaaaaa aaaaaaaaaa aaatgagcgg cc 3892

```

```

<210> 12
<211> 482
<212> DNA
<213> Homo sapien

```

```

<400> 12
gatgaaaaga agaaagggga ggcagggtt cattgaaact gccagcctct aggtcagtgc 60
acatagatag ttactgcctc aggaggcaca attgctgggt gaatggtacc acagagattg 120
atthagagtc atggctagtg ggggtgttct acatgtcatg tctggaagaa aatcagggtgt 180
gtttttgaga cagtgtgttt tcatgtggag caaacaatca aaaccagtga gtgagagcaa 240
ccccagcatg accatgtttc ctacctgtg ccacactctc tgtgaagagt tgtgtccaca 300
ctttagtTTa ttttaataatt tgatgtaatc tgtatgtatc cctcattTTa ttcagaaaat 360
aagctcagag aatTTTTata tcttaaggcc acagaattaa tgaatgacag gagagcctgg 420
attaatgaat gactagagag ttcaggatat cctgactTTa gtaatgtcta catttcatgg 480
tt 482

```

<210> 13
 <211> 295
 <212> DNA
 <213> Homo sapien

<400> 13
 ctcaaaaaca tttgggtatc cacagtttga cccaggaaaag ttgactgggt agatataatg 60
 gtccatatag aaaatttata gaagcagaat ctgtaaatta tgatgtgaaa aagtcagata 120
 atttttatctt aaatatagtg atcatgataa caataattca gttgaattaa aaattagaac 180
 agattaaactt gaatttggtt tgttctataa aatataaata atggagttaa aatgtcatat 240
 atgagtttgc tctggaaaac acattctcaa aataaataaa atctgatttt atttg 295

<210> 14
 <211> 188
 <212> DNA
 <213> Homo sapien

<400> 14
 gctcgagtac ggtgtgggtat gaaatttcaa ttgttttcaa tgcataaaaa cagatattat 60
 gatatagttc attatactat gtgattttgt tactttacaa tcctattata ctgtgtgatt 120
 ttgttacctt aaaatcctta tgtttatggt aagaatgatt cattcttggg tgagcttggt 180
 ggctcatg 188

<210> 15
 <211> 304
 <212> DNA
 <213> Homo sapien

<400> 15
 ctgctctcta ctgccttata aaatccctgt gtcaagggat gctctcagta tcatttgccc 60
 ttgcacagaa tatccctggg gtttgagggt ctttgaatte tccctctttg tcatctcttt 120
 cgctgccact tctggctgtg gtcactagct tggccatagc acctctcttc tccacttctg 180
 atctgctgct tctaaccttc tatagattgc agctggcttt aaaatagatt gtaaagtgtg 240
 aggcattagg ttctgagaca gcggcagaga gagccatgca aatgtttagg acaaccagc 300
 cttt 304

<210> 16
 <211> 212
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (157)..(157)

13

<223> n=a, c, g or t

<400> 16

```

aacc caatct ttcaaaacat gaaagggggt gatggagaaa accttagctt ggttgtctaa      60
agacatgggt gcaaactcta ggctagctct gccaatcact tactgtgcgg gtttgactca      120
gtcccttccc ctactaggt cccagtttct ccatttntaa aacaagcaat tgtgtacat      180
tgatggttta catcaataaa ggttgaaacg gc                                  212

```

<210> 17

<211> 122

<212> DNA

<213> Homo sapien

<400> 17

```

gtaaaactat tacaaaatac tagaaaaagt tatgtgcatt tttagatgat atattcaaac      60
atatctaaag cgtaataaaa aatctaaaat ctctgcaaaa attaaatata gaatatagta      120
ag                                              122

```

<210> 18

<211> 220

<212> DNA

<213> Homo sapien

<400> 18

```

gcgagggtgc atctaaaatg aagtagcttc atggaaaccc atttgctgat gagaaagcaa      60
tttacaactt gttctattga aactcatat ttggagttca acacccatct ctattaattt      120
cctttctcat atcctcaggg ataactttac agaacacctg aaatgttaac aatttgtaat      180
tcaagaatga atgtatctgg gcactttttc aaatacagga                                  220

```

<210> 19

<211> 442

<212> DNA

<213> Homo sapien

<400> 19

```

attgagtact tgctatattg gggatagatc tgggaacaaa atgaaaccaa attatctaac      60
cccctgagat tacaatatcc tgggggagag agacaataag ccagaaatat acggtgaaaa      120
aaaatagaac agaagaagga gacaggtagt atggagtgga agagctgttg cacgtaagat      180
gatcatagaa agcttcacca agcagggtggc atatggtcag aaacctgaaa gaggcaaaga      240
aactggctat gcgtatatct aggggaagag ccttgaaggc aaagagaagg caagtgcagt      300
ggcctatgaa tgagcatgcg tggcagggtc cagtcacagc tagtggcatg gaaaaggcac      360
ggagtagtag cagatggaac cagtgaataa gcagcaggcc actgatgtag ggccataaag      420

```


14

accattcttag gcttgctgcg gt

442

<210> 20

<211> 872

<212> DNA

<213> Homo sapien

<220>

<221> misc_feature

<222> (188)..(247)

<223> n=a, c, g or t

<400> 20

ggaatcccgg gggcgagcac aattttaaaa gaacagtaca ccagcgtaat agagtaatag 60

ttttttcatg ccttggtgcc agtctgttaa gtctgtctga ttgtactgtg cagaaataca 120

tgaattcctc ccagtaaaat gcaaacccaa accttggtaa tgaaggtaaa gtttggggga 180

gagaactnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240

nnnnnnngtc ttgggtcagg atctgggaca cagtcagtac tgtgtgtttt gtgtttgtcc 300

ttcctcatgg taacctggat tccttgatct cccggtaggt ccactccttc atcacacctg 360

ggactactct catatttggt cctcctatc ctgagtccta gtgggttttc atttgactc 420

caggttcagc tgcagacca ccaggtgaag acccaccagg tatgtgtttg ggcagtgga 480

tcacttggtc tggaccccag atcttttctt ctgcttgga aaggtttact agttcggctt 540

catccaggtg tggttccagg ggaattgatg gcctgctgac ctccacgttc tctttccag 600

cacaccttgc cctgcttctg ggccacgtct ctccagtga cctgcaggag acctctgtgg 660

atgcaccgtg tctcctaaca ctgtctccag ccacactga gcttgtcttg cgtggaaatc 720

tttgtctctg ctgttgctt tgcttgagc gtccttgccc cacacctct cacagctgcc 780

tctctgtcat ctttccaatg tctgtcctca gagagccctt cctggccact ccatctaaag 840

gagtcctggg ccaggtgtgg tggccacac ct 872

<210> 21

<211> 438

<212> DNA

<213> Homo sapien

<400> 21

aagaattcgg cacgaggggt gaagaccac cagcacacct tgcctgctt ctgggccacg 60

tctctccagt gcacctgcag gagacctctg tggatgcacc gtgtctccta aactgtctc 120

cagccacac tgagcttgtc ttgcgtggaa atctttgtct ctgctgttgc ctttgctgg 180

agcgtccttg cccacaccc tctcacagct gcctctctgt catctttcca atgtctgtcc 240

tcagagagcc ctctctggcc actccatcta aaggagtctt gggccagggtg tggtgccca 300

cacctgtaat ctcagaactt tgggagacgg aggtgggaag atcacttgag gccaggggtt 360
 caagaccaac ctgggcaaca tactgaaacc cccatgccta taacaataaa aaaaaaaaaa 420
 aaaaaaaaaat tggcggcc 438

<210> 22
 <211> 1002
 <212> DNA
 <213> Homo sapien

<400> 22
 gtggtgtgtg atttgtatgc ttgatttaca ttgctggatt tcctagtgc accttcttgg 60
 ccctccgaaa ttctttgtca tgatagatta ttctttttaa tgtgaccgca tttgataggc 120
 taatatttta ctttgatct ttgcatgaag tactcagtgc aagttaactg ttaaagttat 180
 aaaaccataa gggaagcaca gggctgtgga ggttggggag ggtttttttt gaaggaagtg 240
 ggcattgtgag tagggatttg aagaatggta agactgagat cagggggttag aaacttaagg 300
 acctacaggg gccagagagg tcatatgaat gaatgccaca ggctcagtgg gagacatctg 360
 agaggggttg gaaccggggg gccccggggg aacacacacc ctgcctcaag ggggcagtgc 420
 cactcaaccc cagccaatca ctgccatgtg ggaatatggc cctggtgttg ccacacgggc 480
 tgatcattaa gaaaatccag ctgtctgaat ttttatgtga aacttttcca atttaaaaaa 540
 ttattgttaa ccaattaaaa tttttagaaa acaccttctt tgggtgactc agacgactgg 600
 atttggcctg ttagtacact ggtctgtgtt tgctgattta gacctttaag attctgatgg 660
 tgggtggttg agagaagcag tatgttagtg ataaaaggac acacgagtga gtatcatatg 720
 gactcaacat tgtgcctaact actttccatg cattctcttg tagaatcctc acaactgcct 780
 taggtattat tccgctatcg catagatgag gaggcaggcc cagcaaatta gagtcactag 840
 ccagccagg gagagctgga tgcaggctgc atccaaggca gagagtgcag gccagcaaaa 900
 ggctcattgc caggaagctg tgcacagtgt atttcagtct ggtgggagag caatgtggga 960
 ggcacaggaa gggtaaagtaa agatgtgggc tgggatggga gg 1002

<210> 23
 <211> 906
 <212> DNA
 <213> Homo sapien

<400> 23
 aatgtccttc cgggtgagag tgaagagggc tttgggggga accacaacat ccagggcgac 60
 agcttgagaa acatctctca ggcagaggtg gcatcagga catcccatct ggaactgaac 120
 tagaaagggt ggctatttcc cagtgtgtct cttcctcttt tctgctcccc ctgtccatgg 180

16

cagtaggggt tgggcagtca gcctggacag ctcaaaacaa actactgggc acaacttctc 240
 agaggcaagg aataataggg gtaatgagga ctggacatct ggtgactggg gactcttttg 300
 cgatgtctgt gggaaaggggt taggggtgaag atagaaacag gttagtatgt tcacaactag 360
 ggctagaaaa aaaaaacaga tgagtaacca gtcacatccc tttttctgga aaaatcaggc 420
 tccccctccc agttcaaaag acacatgccc tggtaacctt tgaagtttcc ctacctatgt 480
 acaataagcc cttcataaca gtcagagaag atgggtggcc caactctggt ggaccccaga 540
 actctgagag attccaggaa ttaggcaa atcaggattagg tcaggaagca aggattgcct 600
 ctccatgaaa cctacacaaa tctccatgcc atggaaaagt aattagtaag tagatgagct 660
 catttatgaa gtcacctctc acgggtgctg atgcatatta agcacgtaat tgatgctgctc 720
 tgttactata acgattatgt atttctgttg ctgccacact aaaatgaata acaacatact 780
 ttcccaccta aaacccaaaa gaagaaacca gtgggagggg tgtcttcagc cagcccacca 840
 gcatagaaaa gggtcacctg ccagttaccc caactcccag cggcccaatc ctaggctcct 900
 ccatga 906

<210> 24
 <211> 287
 <212> DNA
 <213> Homo sapien

<400> 24
 gtccgcggac gcgttgggct ggactgagag taagaagagg agagttccag taagtttgac 60
 tacagagcac tcattatcat tgtgattatt atgttcccgc atattttgtc gttcttagtt 120
 aggcatactc ctttaacact attgtaataa aagcaaaatc tggaaatctt taagagtgt 180
 gcagttgctt ccaagggagg ctgaatttta ggaatgaagt tacatataac attgacatat 240
 ctgggggaaa agactagaaa aatatagttt cattgctttt tgtgtgt 287

<210> 25
 <211> 1378
 <212> DNA
 <213> Homo sapien

<400> 25
 gcagagagaa gaaagcagac aggaaacgaa tcccttcctc ctctcccctt taccctctc 60
 cttcactccc tagaccaagc tggtcacccg gctgggaatc aagatttgtt ggctgggaag 120
 acaccagtct ctgtgctcgg gttcagccca gtcgggcaag aaagtctggg cagctggaga 180
 aggtcgtcac agtctcccggt ggaatcttgt tgctgagag ccccataacg gaaagggcta 240
 agacgactct tccacacaca ggtcgcagcc cctgcccggt aggatcccca ggaagatgct 300
 ggggacgctc tgagcatcgg cctttcttta cccccgactc cctggagctt ggtctcggga 360

```

tgccaacttg gggcacggag gcgacgggct gctccgaagc tggaggggtt ctgcttgggt 420
cagaggggatc acgacctcag cagagccagg agaggttggg tgacaccac tgcgtattgc 480
acttgactca aaactgccct cggcctgcgg aagaggcgag ggcaagaagc ctggtggcgg 540
gtcaaacctt caccctatta ttttctatca ttttttatgc ccatatattt gttaaagtta 600
tggctgctat acaccaacag cttattctac aagaagtgcc cccgaggagg attgggttaa 660
gctttgcaaa tttggcttcc caggtaatgc gcttcattat tctgctcccg acttaccac 720
cacaccagtg ggtacccgga gcagcaccca cttggcagaa ctgatgactg cttgggcccc 780
gcggagtgcg cattgcgcta acacgcgcac ggggaattgca cccttgccgg agcctccgca 840
ccgtgcgccc ttcaaagagc tggcgacccc gctcacgtgt aagcaacctc ccactttgaa 900
actaatcgc acccgggtct ttcaccccaa aggactttgc tgcggacgct gctctgacct 960
aagacgcggg agagaagtcc caaaggctac agccaggggc tgggggactc ctctgctcac 1020
cttagtctct gatttcgaag gcccgaatta actatctgtt gtgtcagaat cggcgagtgc 1080
aaagtagctg cgcgccgctg acgcgctctc tcttgggtcc gggatatcca gggcatccgg 1140
agctggggag ggttgtgcgg actaccaag gcacgcgcag ataccgggca gggaagaagg 1200
gctccgtgtt ccaaggatgg tcgtttacct ctctgagat cagccctaata cccttggtta 1260
cttaagcgcc tctgaactcg gggagcgctt ggtggggcgc ctccctgcat ccaccctac 1320
cctggctcct tcccaggccc agcatcttct ccaccattct caccocacc accccac 1378

```

```

<210> 26
<211> 834
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (817)..(817)
<223> n=a, c, g or t

```

```

<400> 26
aaacaatagt atccatcttc ccacaattac tattaaagat tcttaaaatt tctggaaata 60
ttgaaggaaa cgaattatgc tttaagattt taaacctacg gagtaagtta attctgtctt 120
tgccatttac tagctgtgta aacttgga gaacagcttg agactctctg agcctcattt 180
ttttaatctg tcaaattgggt ctcagaacaa ttgtgtagga ttaaaatgaa atattaccct 240
gtaaatagtt tagaacagtt cttggaacat agtaaaaact cagaaagtgt tcagtaaaat 300
gttgagtata ttatttgtat gtgggaattg ctgatccagt tgactcaatt cacttgcaa 360
attactggat ttgaataagg agatgaagat tgagtcacat tgttctcttt gatccatttt 420

```

| | |
|--|-----|
| cttttagaca acatcactca aacaacccta ttacttcagt agcaatgaac tgtctctgga | 480 |
| ttttattaag catctcacta gttccttttc ttcagctgta tggcaccctg tcttcctgta | 540 |
| caccagaggc tcctcagctg ggtaagggtga gccaacgtta ccaggagtat atgctgagag | 600 |
| gccatttcaa agtctttcat agaaggctgt gcttgggcaa gtagaacttt tcatcataca | 660 |
| gtcccagaga tgtgaagtta tcaaggtcag agaagaggaa aagagactca gagaaactgt | 720 |
| gttctctagt ctttcattcg ggaacaaatg gtgtcgtacc aaatggctga aaagatccca | 780 |
| ggcttactct gagaaacatc tctctcttct gaattanact tgcacagtgg ctac | 834 |

<210> 27

<211> 1212

<212> DNA

<213> Homo sapien

<400> 27

| | |
|--|------|
| gcctccaagg agcgggcggc tgggagaggt tgtgccacc gccaccaagc catcggcgct | 60 |
| gcttcggact cccggtaggg ggagccgccg cgggcgccag gccggattgt gttctggcct | 120 |
| cggggcggcc ctgctgccag cggccgggac tctcagcaca gagctcgga ggactccacc | 180 |
| tgctctccag ataaggcgcc atgagcagag aagggaacag acggccaagc tccctcctaa | 240 |
| ttccccgttg ttgcagagca aagaagatgg gggagaacta attatattct ctaggtctaa | 300 |
| atattggtga aaaatttgta ggtgagatca cctcctttgc ccatatccat atataatata | 360 |
| tccattacat gtgtttgccc atatatatac gtatgccata tgtaaatatg tacgagtgtg | 420 |
| tatgtatgtg tatagcagtg gtttagaatt agtctcagtt ccacaaattt tgagcccttt | 480 |
| taatatttta ggtgacctaa taccttctca ctgtattttc caagtcactt atttccaaga | 540 |
| gttttaggtg atcccagcat aggagctgag cataggagct cagatgagac ttgccagttt | 600 |
| tgctgcagta atgtcagcaa aagagtctct ctctttctga ctctatagac ttaagtaagg | 660 |
| gtttacaggc tttcatttta aggtatgact gatattttaga gaacaagact aaaccaaggt | 720 |
| caaggggaaa acagttagac tgtgaacctg cagaccactc cactgcaccc agtgtggccg | 780 |
| tacacctgtg ctgaacggcc agagacttcc tggccaatac tcccacagat aaaagactgc | 840 |
| aggatctgct tatacacaaa tccatctttc gctgggaacc cacaactcat aacaactctc | 900 |
| ctaatagcaa gtcataaggg cctcttgaaa gagaatcata aaaaacaaag aatttcaaac | 960 |
| ttggaaggaa acctaaagtt tactggtgag gagcttgaac tgagtagaag caaagaatca | 1020 |
| tactccagat gagaaacgcc ccatgcactg agctgagact ggaactctcc tcttcttact | 1080 |
| ctcagtcag taagtttgac tacagagcac tcattatcat tgtgattatt atgttccgcg | 1140 |
| atattttgtc gttcttagtt aggcatactc ctttaacact attgtaataa aagcaaaatc | 1200 |

tggaagatct tt 1212

<210> 28
<211> 433
<212> DNA
<213> Homo sapien

<400> 28
aaaaggcccc ctccggcctt cagggggtgg aaggggacctc ccccccccca cccaatccat 60
gcaaattaaa ccagttgttc ccaagaagaa gggatcttgt cagactgact caagtatctg 120
acctaacaga tgtagctttg gtagatatgg ctgcagcctg atgatgtctg gaatcatggt 180
tatctaaaat gaattccctt tgttttgact cagctcaaca agcaaacttt tgattataat 240
ttttggacag taaataatta ttttgttaaa cttgaagatt ttcttcattc accctccttc 300
attctgatgc acatattatg catcagagag ggcagtttca aataagccct cctaaaggct 360
ctgaaatgag ttagaatagc acatttatta aatcaactta tagacaacac taacaactgc 420
ctctacacaa aaa 433

<210> 29
<211> 1359
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (729)..(864)
<223> n=a, c, g or t

<220>
<221> misc_feature
<222> (1030)..(1053)
<223> n=a, c, g or t

<400> 29
ttcaagtgat aacactaaag cattggggca aatttaaata tcaaaaaatg ttgtaattac 60
aagtttcttc ttgtgcaccg agcttacaga ccgagacata aacagtaaga ctaatgtctc 120
tggggagaaa aatatacaat tatcttctat attattggat aaataaatta ctgtatggta 180
atltgtttta atatttgtgt gcattcactg tactgtgcta gttaaaacgg aaaaagtggg 240
ttggtgcatt ttgaaatatg aaagcattta aattttttta gcacataatt ttagagaaca 300
acttagccct accctattta tgtaagaaaa ttgagacctg ttcaggcaga cctgctgcag 360
gtgatatggt aacaaagtca gagttggcgt gtgcagcaca tttatttcac taaaattcat 420
tttgtcatcc acatcatatc aatacaactc ttttaatttta tgtaaactat cttaaataca 480

20

```

aattctgaaa gagccacaat gcacaaaagt tcaaatatta ctttttctct actctctgct 540
aattttaact cgtaagtctg gaaaatacag tccttagttt acaaaaacaa .cttcagaatg 600
ttcctaccta caaaccaagg gtgaggggta cgttaagatc actgctcttt ctttcccaag 660
tgaagatgga aggaagctca cgtctattga gtgccccct tgacaagtca ctttgtgtat 720
atcaacttnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 780
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 840
nnnnnnnnnn nnnnnnnnnn nnnnttctaa tgacttttca ctacacagcc ctgtttgagt 900
tgagaacatc caggtgagtg actcagtcta aatgaaatct gctacagggg aaaaagggtc 960
caggtagttt acatgggtag ctattgtgga gatgtagaga ataacctaga gagatagggtg 1020
acttctcacn nnnnnnnnnn nnnnnnnnnn nnnagcatga agccctcttc tcaccaagga 1080
ggcctgtaga gatgatgtga tcctggcatg aactaggaat gctgttcaca atagagacag 1140
tgcttaagta gtcaataagt aaagtttaaa aagttagtaa aactgagtga agttgatgtt 1200
agtttccaaa tacagtgggt taaataggct tttgtgagtt ggcctgaaaa acagcattta 1260
tggcattgta tgggaagatt atttgagatt tctaaccagc tgtcttacia ataaaacaaa 1320
tgtctaactg ttgaaacaaa attacagtac ttggaaaaa 1359

```

<210> 30

<211> 3705

<212> DNA

<213> Homo sapien

<400> 30

```

ccctatcgga aatttgtttt tagtggtgtt attatgtctt ccttctcacc aagatgaaaa 60
tcaacagagt ggtgatttct gcagtttact tctggatcag ggttcttact gcacagcttg 120
gcacatcata agtgcctgt aactaaggag actaaataat ttccctcca aaggggaaca 180
ccttgggtcg tgtaaggggt agtattaatc ttaatgccag gacaacagaa atgaaatgag 240
aaggacagaa atgaaggctt tatgtttgta tctgaaacag ctttatgaga ctagagtctt 300
ccaattgatt tagagcaaga aaggatttca aatcttagga ggggcttaat catctttctc 360
agtatgtcat caaaagattc tctcatggtg taactctagt taattgagct tattagaaac 420
agaaagcaac atggtgtttc tgaggttttt cttcatgagc aggaggtgtt caggtgaagg 480
tgggatctgg gtgtggcagt gagtatthaa ggtggcgagt gcacgatttc tttggaaacc 540
agagagtcaa ggagctatga aatgacgtt gtgtcctgaa gcacagcaga aagagataca 600
gagaagcgtg aggtcttggt gaacagggaa gtctgggagc ccggtgtcct gttggggatg 660
ggttaggagg caccagcatc tgtgtagctg acagtggcag ggaagagaag agaggggtgt 720

```

| | |
|--|------|
| cctctgagga ggatgaggag ggacaaggtc tggaaactgac cctgggtcat gaccttgagc | 780 |
| atttagtcct tgactcgcat cagcccgatg aggtcccagg acccagtgtc aatgcctgga | 840 |
| gctcctgcca gcagcatcag ggtcacacgg agaaatgcag actctagggc ccaccccaga | 900 |
| cttacttggg ttcagggtcca catatctgac ttcagcaaac cctgcagggg ctgaggatgt | 960 |
| acactgccat ttgagacaca ttttcccaga aagagaggct ggtgggaaga ttccactgag | 1020 |
| gctgagagaa gccccccaca gctgatcttc tctgtctcc atttggttga aatttcacat | 1080 |
| tttcttttct tttgaaaggg gaagctcaaa aggaacatga tgccctgggc tttacagaag | 1140 |
| aaacgagaaa tccacatggc caaggcccat cgggagcag ctgcgaggtc tgetctcccc | 1200 |
| atgagactca ccagctgcat cttccggagg cgggtgacaa ggatcaggtc tcatcctgac | 1260 |
| aaccagggtca gacgcagaaa aggggacgag cacctggaga agccgcagca actctgcgcc | 1320 |
| taccggagac tgcaggccct gcagccctgc agcagccaag gagaagggtc aagtccactg | 1380 |
| catttgagaa gcgtcttaag tatccttgca ccggggacgg ccagtgaatc tctggacaga | 1440 |
| gctggtgctg agcgtgtgag cagcccgctt gagcccaccc ctgggcgggtt tccagctgtg | 1500 |
| gcaggggggc caaccccagg aatgggttgt cagctccac cgccctctc tggccaattg | 1560 |
| gtgactcctg cagatatccg gagacaggcc aggggggtga agaaagccag ggagagactg | 1620 |
| gccaaaggcct tgcaggcaga caggctggcc aggcgggcag aaatgctgac agggagacag | 1680 |
| actatcccta aatacaactg ccagacacag aacaaaactg ttgccgtcat gccagggtacc | 1740 |
| acgtcagtat cttcaaccaa gattggagct cttctggggg gctacagaag cccacggggg | 1800 |
| gctgtcctga gagcgctgag ggataaagga ggagcgctct gcttccggc tgccctgttg | 1860 |
| ctgtcggagt cacaggatgg cggctgtcgt cctgccccca actgccgggc catgctctgg | 1920 |
| tttctctgga ttgctgtggg ttggtgtttg cttggagtag aggtggatgc tttgtgggtg | 1980 |
| actgtgtatg catggggacc tggctgatcg caggcagcct ctacgcgaga aggacacaca | 2040 |
| gagggcggag agctgggtta tgagctcctg aaaagctggc taaagggtt ggtgaccttt | 2100 |
| gaggatgtgg ccgtggagtt caccaggag gagtgggcat tgctggacct tgcccaaagg | 2160 |
| acactgtaca gggacgtgat gctggagaac tgcaggaacc tggcctcact ggggaaccaa | 2220 |
| gttgataaac ctaggctgat ctcccagctg gagcaagaag ataaagtgat gacagaagag | 2280 |
| agaggaattc tctcaggtag ctgtccagat gtggagaatc catttaaagc caaagggtta | 2340 |
| actcctaagc tgcattgtttt tcgaaaagaa caatctagaa atatgaaaat ggagaggaat | 2400 |
| catcttgag caacactcaa cgaatgtaat cagtgtttta aagtcttcag cacaaaatct | 2460 |
| tcccttacac ggcacaggaa gattcatact ggagaaagac cctatggctg cagtgaatgt | 2520 |
| gggaaatcct acagcagtag atcttacctt gctgttcata agagaatcca caatggggag | 2580 |


```

aaaccctatg aatgcaatga ctgtgggaaa accttcagca gcagatctta ccttactgtt 2640
cataagagaa tccacaatgg ggagaaaccc tacgaatgca gtgactgtgg gaaaaccttc 2700
agcaattcct catacctcag accgcacttg agaattcaca ctggagaaaa accgtacaaa 2760
tgtaaccagt gttttcgtga gttccgcact cagtcaatct tcacaaggca caagagagtt 2820
catacggggg agggtcatta tgtatgtaat cagtgtggaa aggctttcgg cacgaggtca 2880
tctcttttct cgcactatag cattcataca ggggagtacc cttacgaatg ccaagattgt 2940
gggagaacct tcaggaggag gtcgaatctg acacagcaca taagaactca tactggagaa 3000
aaaccctaca catgtaatga gtgtgggaaa tcctttacca atagcttttc tcttacaatt 3060
cacaggagaa tacataatgg agagaaatcc tatgagtga gtgattgtgg aaaatccttt 3120
aatgttctct catccgttaa gaaacacatg agaactcaca ctggaaaaaa accctatgaa 3180
tgtaattatt gcgggaaatc cttcacaaca tccactacct ccactacagc atctgctacc 3240
tccactacag catcatggat ccaccacagc aatcaccaca tccactacag catccactac 3300
agaatccact acagtgactc caacacagtg gcctcccact cctgggggta tgcttcagga 3360
aatggaaggg tggaagaagg aggcagagat ccaagggtag tagccagcac tggaaatgat 3420
cctgaggcat cagataatga tgaagataat gatggtgatg ataagaatgg ccatgatggc 3480
gatgatagtg atggtaatga ggggtgatgg gatgagatgg tgatgatgat gatgatactg 3540
acgatgatgg tgatgacggg gataatgatg atggtgatgc tgatgacagt gatggagatg 3600
gctaatttaa caaccccata ctgcggaact ctgggccaga gtttgaggga atgtgaagaa 3660
ttcagacgca actttcatgc agcctctgga aagctcccag gataa 3705

```

```

<210> 31
<211> 800
<212> DNA
<213> Homo sapien

```

```

<400> 31
aaagattctc tcatggtata cctctagtta attgagctta ttagaaacag aaagcaacat 60
gggttttctg aggtttttct tcatgagcag gaggtgttca ggggaagctc aaaagaaaca 120
tgatgccctg ggctttacag aagaaacgag aaatccacat ggccaaggcc catcggagac 180
gagctgcgag gtctgctctc cccatgagac tcaccagctg catcttccgg aggccggtga 240
caaggatcag gtctcatcct gacaaccagg tcagacgcag aaaaggggac gagcacctgg 300
agaagccgca gcaactctgc gcctaccgga gactgcaggg cctgcagccc tgcagcagcc 360
aaggagaagg ttcaagtcca ctgcatttgg agagcgtctt aagtatcctt gcaccgggga 420
cggccgggtga atctctggac agagctggtg ctgagcgtgt gcgcagcccc cttgagccca 480

```

23

```

ccccctgggcg gtttccagct gtggcagggg ggccaacccc aggaatgggt tgtcagctcc 540
caccgccccct ctctggccaa ttggtgactc ctgcagatat ccggagacag gccaggaggg 600
tgaagaaagc caggagaga ctggccaagg ccttgcaggc agacaggctg gccaggcagg 660
cagaaatgct gacatgtaga tgaagcgag tcttgggctt tcgggtccctt tcttttaatg 720
cccatcctca ttctactct gaattgtcac acttttccct tccccaccag ttctttaata 780
aaagtatttg aaaggcaaaa 800

```

```

<210> 32
<211> 74
<212> DNA
<213> Homo sapien

```

```

<400> 32
gtgggggttaa ttttaatatc tatatttttc aaaatgcagc tttctcatct ataattttaa 60
tattacttaa atgt 74

```

```

<210> 33
<211> 1943
<212> DNA
<213> Homo sapien

```

```

<400> 33
attgcaatgg gaatcttaaa ggaaaggctg agaggagcga ggtgaggctg ggtgggtgtg 60
ctgggtgggg tgaggctggt gttctcaggg ttgaacctca tggggtagat caatcagaac 120
ccaggagggtc tgaggactgt gggacctcca agcaactacc ctggactact cgtccatgat 180
tttcagatga gccacactct cttctgcagg aaggagacaa aaatttccga gatggcagag 240
aaaaatagtc accctggtgt gggagtgcc tgaggctgcc agagaaaatc taccaacaca 300
ctaatacttg tgtgacttta cataacatcg agaaagtat tttctcatat actcactggg 360
cggtcacctac gcatgcctgt acaatctaac gcgtgttcag ctagggtctc tgtttgccct 420
gcgaaggctc caggggcagc tccttcctgt cctctgccat gctcctggtg tgcattcagg 480
cactcctggg cttgggcaac ttcactccaa tctctggtct ctgtgtcttc ctcttctgtc 540
tcttataagg gcactgtcat agggactcag tgaacctctc caaacaggat gatctcatgt 600
tgagatgcct aactgataat atctacaaag accccttttg caaagaagac ctcatcaca 660
ggctcagggg ttaggatgtg accatctctt cttggtgaag agtagtgagg gctaccattc 720
agccactac acccagcagt agccattcca tgccctgcag ctgcactgag atgacagaag 780
ttgttgtagg aaattcttcc attttctctt cctggaaagc aaaactaatg caggaagacg 840
ggaggggaga gtaggatgtg tgtgtggagt atgggttttg ctgctgcagt gcagatccag 900

```

24

```

aatagcacag gtctgaatga gagggaaatt aagtcttgct cattcactgc actccagggtg    960
tgtcgtgtgc agggctgagt ggtagtcccc tgtgccagag atccggaatc tcttgctggt    1020
tcctctgccca accataggct gctgtaccog tctgcatggc tcacctccaa gtctctcttc    1080
tggccgaggg gagggctgtc cacgactaat ggtgagggtg gtgcacggtg tcatggctga    1140
gagcctcttc ctggtgctgt gtatcctaac tgcctgtact ctattgggtg ttcaatgttt    1200
gggctgccac ccctgggata tccctccctt tgaaggcatg acccagaagt tatacccgtc    1260
ttttctctct ccactcacac cctattggcc agtatttggt cctatgacca cagctagctg    1320
caagggagtc tgggaaatgt attctttatt ggagatggcc atttgctggg cagaaaattc    1380
taataatgga tattggggaa atggataact gggatgtagc taacattctc tgcaaaggag    1440
gggtctggct ttctgggacc acgctcagca tctcctggct cttccactga cggggagatt    1500
gctggagaaa ggotgatggg tctttgcttg gaaaagccct agatcttatg gtggcctggt    1560
tccttgagga tagcagttat gggacattgc tggcttcact gcttggtttt acaccagctg    1620
atgcccactg cccagcatg acaagggttc agctttctgt cctttgagat gatagccac    1680
tctctgcccg tataggcgaa gagaaatgca ttgactatac actttggtgc taggtactgt    1740
agcaagcatt tctccatgaa ttattgctgt ttgagaagct gcaggaactc tctggcaaac    1800
tgccctctgc aaaataggat gtaaacaatc cccagagta gcctgctcaa ggcttttggg    1860
gacctggggg cagccccag catgacttgg ctctggggtg gtatgttgcc cagcaggtg    1920
tgtgagggtta ccagccatag gaa                                1943

```

<210> 34
 <211> 1052
 <212> DNA
 <213> Homo sapien

```

<400> 34
tatcagctcg gaccttaatt cccacgatca cagattcaat gcaatcccta tcaaatttca    60
ttgacatttt tcaactgaaat aatgaaata atcctaaaat tcgtatgagg gagtaacacc    120
tgctccacag tggcacaaaag cactgtcagg tgccacacac agcctggttg gacatgtggg    180
ccacagtact tccccaccta aagtactgta gagggccagg agacatgagc cagggccatt    240
ttaactgtca gtaatggcca caaagacccc taaatagcct aagcaatctt aagcaaaaag    300
aacaagctg gagacaatac actacctaac ttcaaaatat actataaagc tataataatc    360
aaaacagcat ggtattggca taaaaacaaa cagaccaatg aaacagaata gagagcccgg    420
aagtgaatcc atgcatttat ggtcaactga tttttgacag aaatgccaaag aaacacaatg    480
tggaacgac agtctgttca ataaatgatg ttggggccag gtgcagtgcc tcatgcttat    540

```

25

```

aatcctagca ctttgggatg ccgaggtgga aggaccactt gaggcctccc tccaggctcc 600
ggaagagacc tcctccatga tcccttgtct aaggggaagg ttcttcagga ccttaccgtg 660
ggggctgaag gtggccaccc ctccaggaac ttatgcccc ggcgctgaat ttgggctgcc 720
taagtctgtg tgcgtgcgtc tgtgtttgtg tgcattgtct catgtctgtg tgtttgcatg 780
catgtctgtg tgtctgtgtg gtctatgtgt ctgtgtgtac acttctgtat gtctttctct 840
gtgcattttt gcatgtgtct ccattgtgtct ctgtgcatgt ctgtgtgtct atatgtctgt 900
gtctttgtat ctgtgaatct gtgtctgtgt gtctttgcgt gtctgggtgt tgtctgtgtg 960
tgtgtgtcat gctgttgtgt gtctgtgtct ctgtgtgtct gtgtctgtgt atgtaatggt 1020
gcgtctctgt ggcaagaaga ggggtgtgtg ga 1052

```

```

<210> 35
<211> 1451
<212> DNA
<213> Homo sapien

```

```

<400> 35
ctgtatgtaa tgttgcttcc tatctcctca ctggaatatg agcttcattg atgatgtctg 60
ttctccctag gtatctcgac tagggtttcc tgagagaagg agccctgggg atgggtagaa 120
tgattcacag agacagacat ggcaaggaat gagtgtgttg atgggcagcc cggtcaccta 180
ggtatagacc ctttgaaacc ttgcacaacc attccttaga caatgtaggc caagggata 240
tggagggcat ttcaatcttt tctgtccaaa aagcaacccc caaattagtg gcatctaaac 300
agaagccttt atttttcact tgtgagtcag atgtgagtg cctgtctgat ccagctaggt 360
ttgcttctca tcaccgggtca gctggggtga tagcaattcg gcttgccctag acagctggtc 420
ttcaagctgt ggggtccagcc ggtgtagatc ttcagatctg ctctaagtgt ctaattcttg 480
ggtgcaggat gaagtgattt agctacttgg tggagagcag aggaagatct catggcagtg 540
atggaggggc aggaaggcca ggccattgc acatgcatat ttacgtcata tctaccaaca 600
gccattttgt gaaaacaaat cacatggcca agcctatagt cagagggaaa tacactgtct 660
tttttagagc aaactacaga ttacacagc aaaaggcatg gatacaggaa agagtgaagg 720
gttcaataat tcaactctacc aacagaagga tttcaaggct cagacaaaaa aacagatgcc 780
aaggacatgt acatcacaca gatataaac atgtaaaata ataaacttaa ctataatgaa 840
atgaagcaaa gggaaaactt gcgaagaggt gcctgaggag gtgaggagaa taggtcatta 900
ttctttatgt aaaaagtgtc tgagctgaat ttgaattgaa tgaggacatt ctaacctcat 960
atgctatgtt ctagcaatct tgaaatactt gcaacgttgc tgatgcatgg tgcctctctga 1020
ctctttatgc cttcaaacac actgctatct cttaaagtat tctgtcatca ttcaatccct 1080

```

26

```

atgtatccag ttaattggct aattaacttt atcttccaga tggcctccta caggaagtcc 1140
tccctagctt tttaaatac tctaggtacc cccattctct atacacctct agcatcctgt 1200
aaatatccta ggacctatt cgctaattat ctgtgtacat aatcaagacc gtgggtttat 1260
taatggcagg gatcaggctt taattttctc tctattccta gggcttacag cagtagtgca 1320
cagtagacaa acaattattg gctgaataag taaatatctg aaaggctttc agcatgaact 1380
tatggaatca aaagaggcta aaggaaaatg tagaaggaga agtttacaag ttattttttg 1440
gtacagggtt t 1451

```

```

<210> 36
<211> 600
<212> DNA
<213> Homo sapien

```

```

<400> 36
gtccaccaag aagccaaatg acaacaggaa aggtgatggg aagatgaaaa caaaggaagg 60
tggacttttg ggtatatgtt atagccatgt atgtatgtct tcttttttct attttctctt 120
gttcttcac ttaactgtcc tcagtctgcc ccacaccaac cctgtgtcac tccagcaca 180
cataagacag agcagaagac cccatccttg agctgggtct cctgggtat gggctgaggt 240
aacatcccac acaccaggac gatcttccct gcctcccatc ggtcacatta agacattttc 300
aaagtgtaat attataaatg gacctacctc taaatatgtg ctttacagtt attttatgag 360
gcactcaatt tatagctaag ggtttttcag tctagtgtca tgaaagagat aaaagggtgt 420
tcacagatta tttaagacat aaggctgggc agggatgagt cagagagtca ttctccatga 480
agtcaccctt ggccaacttt gaaaggaaga atgtttaact gcactttggg cgtaaagac 540
aagcatctgg gaccctcccc ttcctgac cctgccagca ccactcaatc ggccagatat 600

```

```

<210> 37
<211> 1595
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (484)..(547)
<223> n=a, c, g or t

```

```

<400> 37
tagagacaag gtctcactat gttgcccagg ctggctcaaa ctctgggct cgagtgatgc 60
ccttaccttg gcctcccaa gtgctgggat tacaggcata agccactgca tctggccaca 120
tcaacttata tataccttgt gttctaagaa gctagggcaa attataagaa tcatgtacaa 180
cagtgttct cagccttggc tgctattag aatcacttgg ggaacttcta aaactcctga 240

```

```

ttcctggacc tgacctttca gaccaattag aatctctgga ggtggaaccc aggcattcagt 300
atttttcaca gttccccaag tcattccact gaagccaggg tggggaacca tgacttgaca 360
aagtgtctcc taatccctga cccttgagca aaaggtgatt tacaacaggg atttcttaaa 420
ttttcatatg cataccagtc acctggagat cttgttaaaa tacagattct gattcagtaa 480
gttnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 540
nnnnnnngac gcagggtggg gtaggggaca atagcaaggt tttaggtggg cccccaatac 600
aggtctctaa attgccttca tgaagagaag tgcccaaatt ccggtttagt taaaaagact 660
ataactaaag gatgatgggt ttaaaataat agagatttgg aagagttggg gaattactat 720
tactgctttc tgctatagtt gacttgatac cttctttcta acctcatggc ataaatagac 780
acgccttata taggcaagac aattgtatat taatcactgg gttgtgagtg catgcatgga 840
tgagtgtgag tgcattgtgt tgtttgaggg gatctcagtc ttccgtgtgt gtacaaccta 900
taaatgggac acgtggctag ggagaagtaa taggcagtgt agagccagtc tggcctgggt 960
ctcacctga ctggcaaata tttcctcttc tgtgcttcag tttctgtatt aaaagaaagg 1020
tggggacagg ggaggcctaa ctacatgagc cccaagtttt cagctttttt tcagcttgaa 1080
aatctatggg tgaaggtggg aatttgcaaa tctatctctg tggtatttgg catcccaggt 1140
gtacctctgt aagaaaacat ataaaatgta gaatttataa aaccaatcaa tgtagtagtg 1200
atttttctca tgccaaaaac ttagaggtga ttgtttgagt ggatgggtgtg tggaagatca 1260
agaaataaga gagccatttc gagtttagtg tgaataataa atttgatcaa aaactccttg 1320
aaattagaat tctattctga ttgggggaaa tttttatttt acaaaattgt tttcacaact 1380
ttttggagcc caactatgtg agaatgggag actcatatat caatgggaag tgtacaaaaa 1440
gggggtaatt ggggttgcaa attcctcaaa ttaaggggtg gttaaaactt tgtgcgttac 1500
ttaatagggg aaaggtaata ccaggagttt cagggtaaac tggggggggg gagggagaat 1560
ggggcggcct ggggggaaaa gtggaaaaag gcat 1595

```

```

<210> 38
<211> 350
<212> DNA
<213> Homo sapien

```

```

<400> 38
acaaactgca cgcatttacc aaccagccct agaatccctg ccttagtaaa acattgtaac 60
tgaaaggagt gtcagaaccc tcaagtacat atttgaacat gggaaatggg tggtaaatat 120
tagcccaaag gcagaaagtc tattctggag gctattttgt gagtcaaaaa aggcaattgg 180
attcatcacc ccacctggcc ttggactect atgatcattt cctcactacta cctgttacct 240

```

ccgcctccag tcttcttcca gctcaacctt ctttatatct cataggggtg tattaccatc 300
 ttcataagtt agaaaatttt catcatctca agatgttttt tacctagtgt 350

<210> 39
 <211> 863
 <212> DNA
 <213> Homo sapien

<400> 39
 atggtgcgca ggttcttggt gacctccgg attcggcgcg cgtgcggccc gccgcgagtg 60
 agggttttcg tggttcacat cccgcggctc acgggggagt gggcagcgcc aggggcgccc 120
 gccgctgtgg ccctcgtgct gatgctactg aggagccagc gtctagggca gcagccgctt 180
 cctagaagac caggtaggaa aggccctcga aaagtccggg gcgcattcgg cacttgtttt 240
 gtttgtaccg ttgaacctgg acctcccttc acacagcccc tcaatcgtgg gaaactgagg 300
 cgaacagagc ttctaaacct acctcagaag tcagagagtg gaggaccggt gagatacggg 360
 gcacgcaggc agcgacttcc tgaaatgcta acaaggatcg taggatcagt tactgctgcg 420
 aggagcaagc acttgcttct tgggggagtt ttgcagccaa cagggaatg ggctttcttt 480
 gtgagttaga ggtagaggtc cggcggcctg agtgattgaa actgctcggg acaatgctcg 540
 tatgttttagc aaacgacaga actgtagaac tgttcctgag aaatcccaac tgatagtatt 600
 ttagtcatct cagacgacag ttagcacagt ttaaaaatga ggcctacttc ttgaaaaaca 660
 gaatccaagg tagttttgtc ctacattga caaatgttga cacagccagt gtaatttcct 720
 ataaccagga aaactgaaag aatatatgta cagttaaaat atgtacaatg ctaattaaaa 780
 cttgtgtaat aagtctaaaa gtaatttaat gaggcttcac ttttatgacc gtccttgtgg 840
 tatgcttcgc caggaatata tag 863

<210> 40
 <211> 1063
 <212> DNA
 <213> Homo sapien

<400> 40
 ggtatttgtc atttaggagg aaatccacca gactagtgtc gcaccgtatc ttccagtccc 60
 ctctcccat cctcactggg gatgtcttag tttcaaaata ggtctcccca agaggtccat 120
 ttattgagtt tgaatggcaa gtgctttgga aacacctgc cttatcagca ctctataaaa 180
 tgttctcggg tatagcatcc ttttgataat atagttaatt aacaactgat gttacagaga 240
 ttttaattgt gagatcaaaa ttaagcaaag agcaatggaa tagaaagagc acattttag 300
 ggcttaccat ctctgcacat gaaaccaagg ctcaaaatta ggattcaaag agaagtgtaa 360

29

```

ctgttatttc agaaggtgga ttcatacagc aatgagtatc agtttcatca gcatgggtca 420
tgaatacaca ttcttagagt ccagatttc tatctagtga tacatactct atccttaaat 480
taaaaaaac aaaaatagat caaaatttta aaacctgggg tgagagaaga aagagactta 540
agagaatgag aatatgcctc tgtttccatt tactttocct ctgtctgtgt ctgtccaatc 600
ttccacccca agaacaccaa ggaaatcaat gttgtagacc cagattgaat ttaagatttt 660
tgaatggaag gaatgaaagc acttaggtgc attttctctt gttggaaact gattttaaaa 720
ctttttatta cttacgttta gggaaacatt tggggttggt tgcagaaggg tataatattt 780
atgttatttc tttgtttaca aattgaattt tagtgcatat ccctccttct tccaggaagg 840
attttaagct tctaaataat acaaataat tcagttcatt taaaaaatag tctactcat 900
agcaaatggt atccaacat aataaattga atatttggt attgacagtt gccctccttg 960
gcaaagcatg caatatgtgg tgtctacca aggaagctca gtgatgatct acttcccgt 1020
gtgaatgggc tttggaacga aatcaaagga tacattttag cag 1063

```

```

<210> 41
<211> 392
<212> DNA
<213> Homo sapien

```

```

<400> 41
gcacgaggca aaccagataa ggatacaaca aaaaaaaga aaactacagg ccaagatcat 60
tgatgaacat agatgcacaa atcctcaaca aaatactagc caactgaatt caacagcaca 120
tcaaaaagat cattcaccat gatcaagtgg gtttcacccc aggggtgcaag gttgattcta 180
catatgcaaa ccaataagta ttgaaatgat cacatgattt ttgttctcag ttctattaat 240
gttcttgatt ctgttaatgt tctgttcaca ttgcagaaaa agcatttgac aaaataattt 300
cagtagctgc tgaaaaagca tttgataaaa ttcagcatac ctttatgata caaaaaacct 360
tcaataaact ggggtataaaa gaaacatacc aa 392

```

```

<210> 42
<211> 438
<212> DNA
<213> Homo sapien

```

```

<400> 42
gtccgtgtgt catggagcac ctggagtgtt ctgtctggaa tgctggctgg gagccttctc 60
ctggcatttg aacgaggggc agctgtgtcc tctgtttgcc gtgtaaagaa aagaggacag 120
agctcagagg agatgaaccc cagcagaaaag ggggtgcttga ccagcaggag agaagataac 180
caagaggggtc tgtgggtgtc tcttctgagc tacaccagtt tccagggtac ctgggaccat 240
ggataactct cagatcagca acttgtcagt tgatttcaa gctgctgttg gctggactca 300

```


30

gactcagcag ggagcacctg ggcgagccct gtgctgcggg ctggactccg gcccatctcg 360
 ctgattactc ttgcttttgc tccccagtgt gtcctcaaga ggtcagagcc tgctttgttg 420
 tttcttcatt accacggg 438

<210> 43
 <211> 171
 <212> DNA
 <213> Homo sapien

<400> 43
 tgtagagaa ctaaggaaac cgtcctctac ctgccacata agaataaagg aaacaatgga 60
 acagtttccc tactttccct aattaacacg tattcccatt ttgaggcagt gagttgctgg 120
 tacctgcttt ctcttcttt ctccaaatag tactttaaaa gtatcttatt c 171

<210> 44
 <211> 869
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (847)..(847)
 <223> n=a, c, g or t

<400> 44
 atttcacctt gttcaagcac agggatatta cactttccca aaattaaaaa aatgtatatt 60
 tcgtttgatc tgatgcttgt ttggggtttt tgaatggg gttgcacctt gagaaattga 120
 cagtatgtaa agttctgatc ctgggtgatt tctttatagc ttgacatgtg tctatgtaat 180
 gtgtgtatgt ggacatgtat acattggaag gatttttgcc tgcattgctgc tttcatgggt 240
 aaactctttt ttgagatttg cattgcatta aacaaaaatg ttggcattct gtcacataa 300
 gcacatgaaa ataactctgg ctgcatcttt cttgaggggc aatgtttttg attgttggag 360
 tattatatag cttgatgtgg tccagaagtg aagtaatctt gtatcaatga gatagcattg 420
 taatagcaag cccttttagt aaggtttagt gtaaatatct gactggtagg ggagctgggt 480
 tctggaacag ttgtgcctct gtgagattac gaaaatggac gagaaagaaa gggctgaact 540
 ctaattttcc tgaaattggg gcgtgggtgt ggagttgtag ccaatactgg gttgagatgt 600
 gctgcaagga agcctggagc acaatagtcc ctgtccgggtg ccccatgtt gggcactggg 660
 aagcggagtg tgttttgcta aaaggaagaa agtaaagaga agaaatggct aaggcttgta 720
 gtaaacgaac ttaccataa atatcaataa aatcctatct cccaggcata tgtcttcgag 780
 tgctttgatt gtgcatgtca gagcactgca aattttgacc cagtgggttaa gccctttcct 840

agtattnttc gtgtacagga ggcacagct

869

<210> 45

<211> 537

<212> DNA

<213> Homo sapien

<400> 45

| | |
|--|-----|
| aatgggtccat tgctaattgtg atagtaaaat tgcaatatag cacagtaatt gtaaattttac | 60 |
| ctagaagttt ctaggatcat gtgggtacag tactactgca ttttttctta aatcctagt | 120 |
| tgcataaatt cattaggttt gcttttctga ctttgatttt attggattgc taattttgaa | 180 |
| aatgggtatta atccacaagt atattatgat ataaatatgt tattactgtc actttcaccc | 240 |
| tgatttctgt ggatttcaga tgcttttttt ttttaagttt catagtgagg attagcagag | 300 |
| acatttaaga acatggcaga gatgtgtgaa tgagtaaact ctgtcttcag tagcccagca | 360 |
| ttctgtcagc actctccaat acccatcttt ataccctagc atgagaactc gtgtttgtgt | 420 |
| attctgctgt ctttaagggt aaatgacaca tctctcagat aagttaccag ggcagtcttt | 480 |
| ttcatatggg gcattttcta tagcaataac ctgctaacac tctttcttta aaaataa | 537 |

<210> 46

<211> 715

<212> DNA

<213> Homo sapien

<400> 46

| | |
|--|-----|
| gtgaccatct tgctctcccc tgaccccaga ggccctacga tctggcctca gcctcctttg | 60 |
| gaccccttatt tctctctgca cctgcatccg aggcaggagc ctgttcccgg aacacccctc | 120 |
| actcagctgc ctctgcagtg ctgtgcaga cccgcctgc tggaaacccc ttccccccac | 180 |
| cctcactgca ctccccccac ctcatcctg tttcacaccc tttccttagt gctagccct | 240 |
| cccagccttt ccctggaatt ccgctcctgg ctctgcccgc cttcagtgtc tgatctgggc | 300 |
| tgagctgggg atatgaaaca ccgtaataat gatgtgtgtg catttcttgc ttcccaacta | 360 |
| gcctgtatgt ttctgaagg caagaacttt gtcaggttca tctctgttcc tcacagcacc | 420 |
| ctcaacgtcg taggggctaa agacaggttt gtgaaataat ggtaacacaa ggtcaggcag | 480 |
| gtcaacagct tagtaatgtc agctcagcaa tttgatcctg cagtgggcat cacctgcgtc | 540 |
| agatgtggag cactgcgggg aagggtgcttt atgttccaga ggcagcaagg gacatctgag | 600 |
| cccaggcctc agacttggcc attttgatag cacctatggt gggttacatc tcttgagctt | 660 |
| ggtaagagat ttctaaatat cggagaaaag catttagaaa ccatcctgag gttgg | 715 |

<210> 47

<211> 2201

<212> DNA

<213> Homo sapien

<400> 47

| | |
|--|------|
| tagagtacta atatttcttt attgacatca aaaaggggtg ggaagaaaga attaaactct | 60 |
| tcttaatttt ttaatcctat acttaaagaa ttaaaggaat gagtttgacg ttccaagtta | 120 |
| atctgtgttt agtttgagtg tggtacaaca cttggatctg aagctgcac atttttaaac | 180 |
| cctcactgtc actttcttta gcatcacaga tgacacgggt ccctcgctca cgtcatgcaa | 240 |
| gttctcctcc tacccttttc cctctttctg tctaattg cactctctct catgatgcaa | 300 |
| ttcctgtggt ggactctgtg ttccagaaaa taaccattac tgtttacatt tttaaccatt | 360 |
| cttgatattt gtatgcattt ttcaaaagaa aaagtagcac tagttgtcat ctttaaggct | 420 |
| ggggatttga aagtttttgt tttaaaatat tttttgttcc ttttaacaaat ttcgactcag | 480 |
| cagtgaagac aactgcctgg gtctactttc acccctgtaa catattcctt gggtaatgta | 540 |
| ttgaaagtta acgtgaccca tccacattac ttccatttga aatacagagc agttcagccc | 600 |
| ttggggtcac cttctccctc ttttccccac tctgccctta ggcactcca ccctggtctc | 660 |
| tccctgcct atgttcttgt tacctcctgc cagccactcc aaaggcagcc tcagatcaga | 720 |
| attcctcata cctcccttgc tggatgaact aaaatatctc caaagtagca gcaaatatgg | 780 |
| ctgtaaaaat gtccatgcaa aatgcagatg cccttcttgt tgetgtgatc ttctgaaaga | 840 |
| aatgggatca cggcctcctc tgggtctgtt ttgttagccc tggaggggag ggggcaaaca | 900 |
| ccacagacat cttgctagct tctccaaga gccaggaaca tgtcagctct cagcaactca | 960 |
| actcttcctt ttcatagaca tctggtctca gggccaggca tcacacgaat aaaaataaag | 1020 |
| tacagtagag taaaattttg gctaataatg actgttccct actcagagtt ggtaggggtga | 1080 |
| ggatacaaaa agccatagaa aagtattgat actctataat aaattacaaa catatcattt | 1140 |
| aatgttattt ttaaaaatgg agaataaag cagtgaactt agcttaacag tgaaattaat | 1200 |
| gtgcttatct tgacttttgg aaggaaaatt tgaaaggaat ggtaagctg ccaccagta | 1260 |
| taaaatggtc agccataaga ggggtgaat gtagaccag acagaccacc tagccttctt | 1320 |
| tcccatatca cccctttttt agtggtgaaa acaaaacacc aggtaccaa gaaactgcc | 1380 |
| ttcctgttta atgtgcttgt gactttatga tccaagagtc acatatccag gaagaggtac | 1440 |
| ccaaaaatct agtgctgcag aactagtta tagcagaata aactccatgt atccaaaata | 1500 |
| aagcagcatc ttaagtcagc ctaagcacca gtgttgagaa actaggggtc cagagctgcc | 1560 |
| ctttctggga atcctacagg ccctgcgagc atgagacgga cgtgagcagt tgatacat | 1620 |
| gagcagtgat ttgaagcatt actttctggg ctttggttat ggctattacg gtcaggtgaa | 1680 |
| aatcatgggc atatagtgtg aaccacggcg aacactcatg aaacgtaaac acacagttgt | 1740 |

```

tacgaaaagc acattttaaag tattattaga cgtaaatttt ggactgggga ttctcagcga 1800
aacctgtttc caaaaggggg tgacctagca aggggtgccc acatggggta cccgggtacg 1860
ataggaaacc tcacacgaac actgggacca ccgaaacgca ggcgacagga cggaagaagc 1920
gggggtacac gagacaaaga agaaacaccc ggaagagcga ggctacgaca acaggaagat 1980
ggccaacggt aagaaacgca ccaagaagta tcacagatag acaaaagaag atgaaagaaa 2040
gaataagaaa acgcgaaagc acaagtacaa cagtggcagg tccaaccaac aggtaacgaa 2100
aacaagagta gaaacgaatc cagcaaacia caagcaagac cagagcgat catacataag 2160
aggctcctct cgcttgcttc gtttcccgt gtcgtagtgg g 2201

```

```

<210> 48
<211> 522
<212> DNA
<213> Homo sapien

```

```

<220>
<221> misc_feature
<222> (428)..(428)
<223> n=a, c, g or t

```

```

<220>
<221> misc_feature
<222> (479)..(479)
<223> n=a, c, g or t

```

```

<220>
<221> misc_feature
<222> (506)..(506)
<223> n=a, c, g or t

```

```

<400> 48
cgctgaccag attcttgtgt ttcacagaa gacttggttt ttgtactttt ccctatccct 60
ttaatcattt ggttatgatt tagttttttg gaatctttta ttttgaaatt tctttggaat 120
ttgttctttg gaaagttttc aaattatttt gaccaagatt gaatgtaata ccagtagttg 180
cttaactagt gcttgtttgt tttcttggtt ttgctaactc atcaagatac tctcttacct 240
gtgtctttta atactacagt taacacaaag tttatattta ttcatacaga tataaaatca 300
tgctcatgca ctttcttggt tttggatttg tgaaagggaag tctttgggtt tttttttttt 360
gagaacgccc aggtagaatg atttgtaact gttaaacgat gggatcaact aagacaggtc 420
agttttcnaa atgagtaatt gcctgtccag taccaaata ccagtgaggt ttggagtenc 480
aaagttttcc tgagccccct gcaagnccca aatagcagta gc 522

```

<210> 49
 <211> 406
 <212> DNA
 <213> Homo sapien

<400> 49
 gctcgaggaa ccactgcgct ggcaaggatg tatcattgat gcttaacttg cttatcattt 60
 cccactgga ctgtaagatc tatgtgggca gggacaagat tgtgtctgtc ttaatagtgt 120
 ctctaagcc cctagaatag tggccagatt caataggcat ttgataaatg aatagaagaa 180
 ttaataaaca agttaatgag ttccagttat gaaaatattt gagttttaat tttaaacctt 240
 atgtaaagca tttattctat aaagtagagt tgcataataa tttcatgttt tctatattga 300
 acttaatcat gtaccactta aactaaaata gtaaacagt aaatattaat tataacttta 360
 ttattttatc taaaaggatc aattctgtaa aacttaagtt ttcttt 406

<210> 50
 <211> 2898
 <212> DNA
 <213> Homo sapien

<400> 50
 atgggccaag gtaacccta aggggttttc ttaggcttct tggagcttag atttgtatgt 60
 atatcaaaat gtctttaaaa tgttaagttg ggcagaaggc agttgaagtg agctttcaag 120
 gtatgggagg ttttctacat ttataactat ttcaatctat gcctttaaag ttgcttatga 180
 ttttagctgt acactcattt ttaagggga agaagtttc ttggaccatt cgcctttctt 240
 agatgtcttc actccctgtg atctcataaa actgcctatt tgacatctct atctagaaat 300
 ctaattaaag ctcaactca gcatatccaa aactgaattc ttggctcttc ctccaaact 360
 tgcttctcct tcagtcttct ccaactcagt aaatggcaat gccatacttc tggttgctca 420
 ggccaaaac cctgaagtca tccttgatcc ttcttttagc acccatatcc aatccattag 480
 caaatctggt agaccctacc ttcacaatat atctaaagct gaccacctct tttagacctct 540
 actattaacg ccctattcca agccaccatc atctcttccc tggattgaag ctgtcatcta 600
 caaaaattct tcttatatcc ttgcattcct acagtctgtt cccataaag tagccagaat 660
 gattatTTTT aaaacaagtc aactcatacc attcatctgc tcaaaacat tcattggctt 720
 ctcatctcac tcagagcagt caaagtcctt aaaagttgca ggcctagact ccctgtccta 780
 cctcaggtac caccatatcc cacctcctca tgcagctcca ggcaccttg cctcagtgct 840
 cctcaaagca tcaagtatgc acttgccctg gacagtctgt acttggtatt ccctctgcct 900
 tgaatgttgt tctcccagaa aaatgcatag ttcaactctta catccttaag gtgtcaactc 960
 actgttacct gagcaggtcg tccttgaata tatacatcag cattcccttt cccctctgct 1020

35

| | | |
|-----------------------------------|------------------------------------|------|
| ttatgtatgt ccatagcact caccacgac | tgactttact aagtatttat tcattttactg | 1080 |
| tttgttttcc cataccgaaa tataaacttt | ctaaggacag aaatttttgt gctttattgt | 1140 |
| tgaatctcca atttgtagaa aaatgcctac | cttatattaa acactcagta aatgtttatt | 1200 |
| gaacattaaa agtattacta atagaacttt | ggtttttgaa agaaataata actttaatta | 1260 |
| taagacgtat atgatttttg cagttttact | tagtgtgaca ttgggtttat gagaatcgtg | 1320 |
| tacattcaag tccaggaata ataatggta | tccaaattgt ttgaaaggaa aataatccca | 1380 |
| gtggcaaaat gatggtagaa tttgggtaat | cttttttttt tttttatgaa aagagatttt | 1440 |
| attgaaggta aaacattaga ggttcattga | gaatctctaa atccatgttt tgacattgtc | 1500 |
| aagctcattg caacttccag attgagtaac | acttataaca catttccttt tcaaagtga | 1560 |
| agatttttaa aagagacttg tcacatatcc | atttggtctg tttcaaattg tgagctgaat | 1620 |
| gctgggtaat ctctactagc tccttaata | ca gatttaaaat tctcagtgtt tcctagtgt | 1680 |
| ttctgcatac tttatgtgag ttgttatagc | tgtaacatta cactttattt gctgtttgtg | 1740 |
| tttctgtgact tttggtaatt ctggcattta | gaaacctttc actttgcttc aaaacgtagt | 1800 |
| tatatatttg agttttcatt tgatatataa | ttatttattt tgccctttta tttcccaaag | 1860 |
| acattgtaag ggttaattag atcattatat | tttattatta cagattaaag ttgggcagta | 1920 |
| atcttaatta tgatggaatt atcattatgc | taagtaatta actttaccta gtttgtttta | 1980 |
| caactagaac ctgccctaaa tgttgaatat | cttcctagca agaaacagtc tgtcatttta | 2040 |
| cttacacgat gtctaaccac accataaact | tacataaact agtcgtttcg gtcaaataga | 2100 |
| aaaatgtgtg aatgccataa aaacaaaaat | tctcagttta atgatactgg gaaataggga | 2160 |
| agacagcaaa gtgagacttg ggctcaggat | ggttcaggaa aaaaaaaaaa aaaagacccc | 2220 |
| tgagtaccat taatattcct cagaaattat | tatttcaaaa ggaaatattt ctgtattata | 2280 |
| aatttttcat gagcagccat tatgaaatct | cacaagaatc atagaattca ataaaaagg | 2340 |
| tagaaagtaa ttttagatac ttaaactact | ttaaattaac ataaacattt ctaaaatcat | 2400 |
| aagcacataa atagaactta ccaggccaga | aacgcaaaag ctgaaggcaa ccatatctat | 2460 |
| cctgttcaaa aatgtgaaac ccaggatgat | ctctagaatg agatgcgcat gaatagagag | 2520 |
| gagagaaaaa caaaaaaaaaa aacacaaaca | accacaaaaa agaacaaaac aaacaccaa | 2580 |
| ccaagataac accggaacga taaacagcag | aaacagagaa aaaacataag accatataag | 2640 |
| ccgggaaact aaataaacac aaaaacacga | aaagaagcta aaacaaggga aacaagccta | 2700 |
| agaaaaagc gagaacagaa gtagaaaaca | ggaaaaaaag cggaaaaaca caagaaacaa | 2760 |
| caacgagata tagaaaacaa gcgaataaag | agaaaaaac aggaaaaaga acaataaaaa | 2820 |
| aagaagaaaa aaaaaagcac aaaaaaaaaa | aaaaaaaaag aaacaaagaa aaaaaagcag | 2880 |

agaagtatat acaacgaa

2898

<210> 51

<211> 678

<212> DNA

<213> Homo sapien

<400> 51

gttgtggttt gtcagcagtt tacatattaa tattaacaaa agggaaaaga ctaaattatt 60

gtttacagat tatcttattt tacaatgaga aaacctaaaa cattaattaa gaggtcagaa 120

taaagtggct agatacatta aaatcagcat gtaaaactgt gatgaggaaa atggtgcac 180

ccagcttggc tgtgataaga tcttaatgac aggacactgt cagccaatgc ccagagacag 240

cctgggcatg taagccccct gagaagctct caagggttttg aggagaatcc ataagcagtt 300

aaattctgtc ctttgataac cctcattag ccttcctagc agcagactgt ctagactgac 360

ttctttgctg gacatgaacc aataagttaa ccagaaaaa tgtgcaaaca aaattcatc 420

tccttgacta ctttcagttt ttctgtcttc ttctcctac aaattctggt ttttgccttc 480

ttttcttgac cacagtctac tgattttttc ttccttgat gagtaaataat gcatgcatat 540

gacaaactaa agattatgct atatcaataa tcatttggaa aataatttta aaaagagatc 600

acattgtaaa aatgaatcca tcaaaatcta aaaccccccac ctggcatgaa tttcaataaa 660

aggctcttttt aaaaaatg 678

<210> 52

<211> 1107

<212> DNA

<213> Homo sapien

<220>

<221> misc_feature

<222> (764)..(833)

<223> n=a, c, g or t

<400> 52

aatatgaggt gggaaaaaga aatgaaagga gccacatgct ccttcctgcc actggtactt 60

ttcatatgct tttccactt cttctctgt cccccaattt cacatcatta actccccaat 120

catcattatt cattgaccga gtcaaatccc tactatgtgc tcttgtagtc atactttcct 180

cagggtataat tttgcagttt ggggggtggtt atgtaataaa tatttcccac attggactat 240

aagcccctat ggcagggact gtacctattt ctgctcacca tgatattccc aacgtgaccc 300

tgcagtaggc tctccatcaa tatgtgttta attattgaat gaatgtactg agtttggtt 360

tgttcaaagt catactgatt taaggacaat ccataatgaa atccatcact tacaacaatt 420

37

```

catgctaadc atatgattca cctgtgtaac tcattaaaag ttactactga atgtacaaag 480
cagggaagaa aaaaacaaga aatagaaaaa aaggaagaag gctgaactaa agcactaatt 540
ttatagggttt agttttgtca gaatttagga cattttggaat cctaacatta aaaggggaatt 600
tatagatgtc tgttcatacc ttgtacagga attcctttgta cagcatccct gtggaagggc 660
attttaaccc acattcaatt ccttcagtc taagaaccag ctccaaggca gcttgctctt 720
ctagctccgt agtagccacc ctgactacat gtttgaatca cctnnnnnnn nnnnnnnnnn 780
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnggaggca 840
ggcatcatag tgtttacagc tctccagatg attccatgtg cagccagggt agctgctagt 900
tgggtatttta aaaagccctt cttagggttaa tttcttaca agattagacc gtgtctttgt 960
atcctctgca ccaaacagaa gtatagtgtg tcttaaatct ggtgaataaa tgtacatctg 1020
cttcttgtaa tttatgtgtc tggccttagt ccaacattct gaaatgacag attgccact 1080
ccagtaccag cgacagactt tgcaaca 1107

```

```

<210> 53
<211> 816
<212> DNA
<213> Homo sapien

```

```

<400> 53
attctagtat aatacatcat tgaccaaadc tcttccttta gtttttctga ggtagtgcac 60
ccattaaaa aagtctcttc aattagattt ggcttttagt tagaaagggg ttgggtaagt 120
catgtcttcc atatgacatt tctactcatc ccacactgc ctccactta acctgtacag 180
acgagccctt acacatgcac atgtgtatac aactgtatat gcagcagcca gtgggctaac 240
aaatatcagc aaagaacttt tttaaaaga atttgtaata atccgtgctg tctagaaatg 300
tgagttactt accttattag gaaaccggct tataatacta tatatgactc tgctgcatt 360
ttatagttaa gaaatgtaca taaaaattgc tatattttca cagtttcatt ataatgtctg 420
tttagtaacc catgtccact gatactttaa atatgctaac tactgagcaa ttgcaataga 480
aatccagatt ttcataagaa aatgaaatag aaggatactg catctttaaa aaatagagtg 540
aagatatatt actgtattcg taacatatgt gatagtggaa aaagtatttt caaactcaca 600
aattttacag aggttgtaaa tagtttgatg aagtatgttg ggaaagagag cttctagttt 660
tcatcacaat aaaaagaaaa aactgtttca gatatgcctt gtatatctta cacacaatga 720
gtgaaattgc catgggaact ctgtccaact ttgtcagagc atgctcttcg tgttctgttc 780
aaaatagagt ggccacttgt catggtttag ccagga 816

```

```

<210> 54

```


38

<211> 640
 <212> DNA
 <213> Homo sapien

<400> 54
 ttatctgtat gtatatacat atatgtttgt atgtatacat acccacatta acattttatg 60
 accctgtaac atttgttttc tattctttta ggtgatgtga attgacctca gtgaagaaac 120
 atttcattta gtctaagcct tgcattcagg gtcttacaac agatttcttg tgcaagaatt 180
 attattagtt aaaatggaag gtgttagttt aacaaactga gagtcaaatt ttctaagttt 240
 ttaaatatgc attttaacat gcattatttt tggcatttaa tttaaaatag ttttgtttca 300
 tgtatacaac agctgttaca gatatatgcc agcaatatcc atactttttt aacaagcttt 360
 tgttatatat tttctaactc ccataaaagt ctaatcatct atgaaaaaga tatttaggga 420
 gaaaatactt gttttctgta ataaaacaat taggttgggtg ccatgtatgc caagttggat 480
 aaaagaatac aaagaaacag ctgaaggttt tgggagctcg ctctgacatt aattactaga 540
 aagcaaaacta aaatgcaaag tgaatagtta actcagagat tgattttgag agggggccac 600
 ctgtccctag gctaggtgca gtgcatatca cggctcactg 640

<210> 55
 <211> 404
 <212> DNA
 <213> Homo sapien

<400> 55
 cagagaaagc cacatgatta ttaatcgctg tataccatcc cgggtgttctt gtgtgctaac 60
 atgggaccag tcccgtaatg actgctgttt atattgtgtc ccactttcca gtaataaaaa 120
 taacactatg gtgaacatct ttttgtaccc tgtgggatga aagattctca gaagtagaac 180
 tgcagagcca aagaatatgt acattttaaa gatatttgagt aaaagctaca aaactgcca 240
 cagagagtaa taatttgcac caccacagcc atgagatgcc agtttccttc accctcaact 300
 gccttgattt gccacgtag atgctaaaat ctaaacagcc caatcaagaa atacttatta 360
 acagagagat ttgaacataa ttgatttgtg atggtaattt tttt 404

<210> 56
 <211> 1350
 <212> DNA
 <213> Homo sapien

<400> 56
 catccccctt cgtgtgttac aaagatataa gccctattct catggagctt actttgggag 60
 tggtgatact caacagactg taatttgaag tagaagcaat taactttgtc gataggttct 120
 tggagatata accttaaaca aaatgagtat attgcagcct ctttgtgggtc atgcaatggt 180

39

```

ataagtgaaa ccatgttgaa agcagtggca ttatttgagg acctgctgtc tgtegtttca 240
cttaaagtta gttttaaaga atgtattgac cacattaagt gaggatctac tgtaatctag 300
ggatggcccc tggagaagga aggtagcatt taggttgaag tttgtgggga tggggatagg 360
taggatttgg aaagagagag gggcagagag tgcgcagggg agcacttagg atggaggaag 420
tgggaatatg cacagagatg aatggtggag agcaatctgg tttgacttaa gatgttactt 480
acatacgtgt gtaaagtagt aggttaatgg gcttaacttg agaggatatt ttgggaccag 540
attataggca actaaagttt tttatatattt tagatagtaa gaaaccatga aagtgttttg 600
agacagaatg ttgagaaaca tgatgaaaat tgtgatttgg gcaattaatc tggcctttgt 660
gtagagattg attggagaac agacaccatt agcaacatgt tttagtagtt tgggtgctat 720
gccaaaaaca actggaacag taaatggaaa gagtgggacc aattcaaggt aaatggggtg 780
gccaaacatt ttgatgatta gagactatcc tgatttaggt ttgaagcatg gttattatta 840
gtgcccatth ttgcttttaa aagttttctg ctttggacaa taaatcatat ggtgacccta 900
aagatagatt gctgtagcat tgatagcact tgctaattag attgaaagga gacagggatg 960
gtggtgtaag agagcagaat atacggtttt ttagaagtta ttatttcact gagttttcat 1020
atthttgaaga attaagcata aaatttcctt acctaaattc tgtaatcaaa tcttgtatgt 1080
ttaaaattat ttgtgtgcac atgtgcacgc gcgtgtgtgt atgtgtgtgt ttaacgaaca 1140
catttgagtt cttggaatat tcagcaaaat gatagttgct ggctgggtgc agtggctcac 1200
accttgtaaa cttagcactt tgggaggctg aagcaggagg atcacttgaa gccaggagtt 1260
tgagaccagc ctgggctaca aaacgagacc ttgtctctac aaaaaaaaaa aaaaaaagtc 1320
gacgcgcccg cgaatttagt agtagtaggc 1350

```

<210> 57

<211> 824

<212> DNA

<213> Homo sapien

<220>

<221> misc_feature

<222> (184)..(468)

<223> n=a, c, g or t

<400> 57

```

atcatctcag gtaccaaata gacagtatta tttaaattcc ccctttgaca ttttatgtct 60
agtttttcag accatggaaa tctacaccta tgatgctctt ccaagatcac tgggcatttt 120
ttttagtatt agtaagtgac ttatttaatt tgcaattctc aattatttca tttatatgta 180
aacnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 240

```

40

```

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnntg tgagccacca 480
tgcccaacca taaacatttt ctttctttct taaaaagtta gttactcatg cttagtgtg 540
tctagcgata cttggttttg gttggttggt tggctttatt aactaatccc tcacagacct 600
caggggcttc agctattccc aaagaattag ttgtagtaa ttgttactct ttttttcaca 660
cattatttca gacagcaa ataacattgat tgtaagtct ctaaccacta acaagctagc 720
ttaggtatct caattcacct aactacagt acctgacatg ctagcttaac tgaaaaaagt 780
agtatcaatc tgtaatcagc tcttttcccc caagatccta ccct 824

```

```

<210> 58
<211> 832
<212> DNA
<213> Homo sapien

```

```

<400> 58
gagggttctcg gagagagagc tagtttcccg gttgcgctgg ccaaaggccg gcctgaactg 60
gatcctgcgg accgcttctc agcggcttca cccccacacc actcgaatgc acagatcggg 120
acacctcagc tgggtgactt ccctcgttgc aatcggcctc gggagcgggtg cggcgtgtac 180
cgagcggcgc ccctgggctc agcgcgctgc aggcgccccg ggcggatcct tcaccgaggc 240
aggactgggg ctggcgccag agcgggtgcg gagcgggtgtt gggtcaggcg ctggcgacag 300
tgctatggcc acggcagggg ccgcgtgcgt ctacagcgggt gcgccttcgg accctattgg 360
cgcccgggac ttaagcgggg ccgccgctgc tggaccagga cgaactggag aaggagacgc 420
gggcggccac ggtgacctcg ggcgtagggc cggcggcgccg gggcacgctg gtggaagcgc 480
gggcccgcggc aggggaggac atcgggcccgc agagaaagtg ccgatggcct ggggcagcgg 540
tgcgaggggtg ggaaaaacgc aggatattgc atcatagaga cgcggccacc ttcgcccctg 600
gagagcgcca gtctcaggag cgccgcaccg gtcacgggtg gaggtcagcc aggcctccgt 660
aagcgcggtc cccttcagg agtagccttc gcctgggctc ttgagtagat gccaaagta 720
tccgcacgt ctctctagat atctgcgcta aaggccaccc ggtctcctta tcagaggggc 780
aaaactctcc tcgtgggtct tcatacctaa taaaagtcgg catcaaacca ct 832

```

```

<210> 59
<211> 534
<212> DNA
<213> Homo sapien

```

<220>
 <221> misc_feature
 <222> (143)..(169)
 <223> n=a, c, g or t

<400> 59
 gaaaatgagg ttacagagt catttacttg tagttagtca tgatttctat tatattgaag 60
 aaaaacatat ccattgttta gtttacctac tagagaaaat agagttatat ggcttcagtt 120
 cgttaagtta ttacaaaac cannnnnnnn nnnnnnnnnn nnnnnnnnng caaatacaag 180
 tcctggactg gtaggtatga acttaattgg agcaattttc tgtacagttt tccttgagtg 240
 gaggtaacta agtagcttca gtataccttc agattgaatt ccagctcatt gtagcggag 300
 aacatgggtca gctacagcta tgtttcgaga tcttggttat tttaaagaag tttccgctgc 360
 agttgtaaca gatgtggaat tactctagtt agcaaccag ggggggttaa tcgctaaccg 420
 gtctctgtag ccttctgggt acggaatatg gctttccatt ctggaacacc acagaatctt 480
 tcttagagga atcctcttta aagtgccaaa tcgtgatgcc tgtgtccctt tacc 534

<210> 60
 <211> 1780
 <212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (1036)..(1073)
 <223> n=a, c, g or t

<400> 60
 gaccgcttaa ttaaagatct tttttttttt tttttttttg ccattggtag agcaaattcc 60
 agtgggtggt aatcaagaac tctaaagttc agtagagaca ggtgttttga atgtcaagga 120
 aatcactgag gtagatttgg gattacaata agacagctgc cctgtgaggt cataagagct 180
 tttgtgaggc ttactccagg ggttctcagt cttggcatta ctgacatttt gaattggatg 240
 tgttcttgct cctgatgtgt ggggtgtgta ctgtaggatg ttagctgca cccctgaccg 300
 ctatgcataa atgccggtac tgtccccga agttgtgaca accaatgtct ccagacattg 360
 ccacatataa cctgggaaaa aaaattgccc caagttctga gccactgggt agaactaagg 420
 gtagtttagg gacaatgtta atgaaagcct ctccaaagtc ccatagtctt ttaataatt 480
 aattctgcag aatttgctct ttaaaaacat attttttaag gcaagaccag cctattcata 540
 aaaattgcag ttgtttccac ataagaactt tggatattta aagaaaaatg taatttgaaa 600
 tgaaaagccc ttataaatcc ctttatactc atcagtcatt ataacagtaa atgaaatgaa 660
 gaaaaagcat ttgatctcac tcttctgtgc tgatctcttt aactcctgaa tatcctcaat 720

```

ttgatgagcc cactatccaa tgagactgga tagttatcag cacaattgat tcatgggggt 780
gagatgttta aatctccttc atatttaaaa tgatcgattt taaacttgat gatgttaatt 840
cttcttagtt cctggttaagc agctcatctt agtgctaag acaagttcag atgttaaaaa 900
aatctttatc atttatggct aattttatcc ttgccaaat catatctatt ttcttcaaga 960
gagcaaaata catttcctga ttttaaagta ttctattttt aaatatactt tcacattcct 1020
gtagtaaaag actgcnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnctgatgt 1080
agaaaaatg aagaaacatg acctgatttg aattttgagt ttggatgaat gatcgcgga 1140
tgcatgcctg ctctatagaa gtcatgcacc tgataaagtg tcacgtattt gtgcttctta 1200
gtttatttac ccttgagtag ccacaggtag aaatgaggtt gtcaaaccag gaatttagtt 1260
tctcaagtgg aaattgcatt gaaagataat cctcaatgta atctgcatgt tcttaccat 1320
ctggagttaa gataatgcag acgagtcatt ttctgggtgt tctcttttca tgtgttttat 1380
ttcgtacctt ttgttttaca tggaaactttc ttgaatgttt gctagatgat caactgtcac 1440
ctcttagtac aaagccatgc cagacactgc acctactctg cactctaag agaacaatcc 1500
ggaaaggatg attttcaagg gagagtgacc tcttctgga gatctgaggt tatgttacag 1560
tattgtggag ttttgtgtgt taaaattctc ctctgtctct cacaggcaat ttgtctagag 1620
ttgcaatcct cacatttgag agaactgtca aaactctaag acataaagtt taaataactc 1680
attaataatt gcatgtgaat cccccagtaa tgtatgtaca tgagcatacc tttcaaagat 1740
gttcaatggt gctgtaaata aaattttttt aaaaaaaaaa 1780

```

```

<210> 61
<211> 382
<212> DNA
<213> Homo sapien

```

```

<400> 61
tgatgtctta gaaacaagcc ttaaaagatc ttaatcttac tcttgctaaa tgtagtataa 60
gtctaagcca gcctcagctc ttggcctgag attactagtc tcttggtttc tattctacat 120
gtattctcta cacagcagtg agggtaatca ttgcaagtaa aatattgtct tacttatttg 180
cttaaactct tcccatagtt tccctttaca cttagagtaa aatccagacc ctttctcctg 240
atctgtaaga ttgtatgcag tctcttgctt ccttagttct tcacccatgt tacctactgg 300
tactctactt gtctcctgat ttagctacac cagcatcctt gataaattat tcaactgctt 360
atcccttctt atcattcagt gc 382

```

```

<210> 62
<211> 969

```

43

<212> DNA
 <213> Homo sapien

<220>
 <221> misc_feature
 <222> (949)..(949)
 <223> n=a, c, g or t

<400> 62
 cctttctcct gatctgtaag attgtatgca gtctcttgcc tccctagttc ttcacccatg 60
 ttacctactg gtatccact tgtctcctga tttagctaca ccagcatcct tgataaatta 120
 ttcaaaaagc caagctcatt cctcatggcc ttttagaatt ggattataaa gaggggtgaac 180
 tgcttatccc ttcttatcat tcagtgtctgc tcaaaaagtta tcttctcagg gaagattttc 240
 ctcaccattht tatctaaact atggctcttc tctcccaaat cactgcctat cctgtatgct 300
 gcttttaatt tcttcttagc atatatctga aattatatta tgtatttgct aatggctctt 360
 tccctattag aatgtaagct ctatgagggc aaggactcct gtcttggtta ctgctgtatt 420
 cttctagcat aaacacacac acccccttag aacaattctg gatacacaat agaaattcag 480
 caaatgtttg ggtgaatgaa atggccctaa aatactattht taaaacttgt tttctttcca 540
 ggttatattht tcttatthta tgtgtgtaaa aatgtgggtg tatgaagtht tttggtttta 600
 aaaccttcaa tagtgagtht ttgtgggcac attgtattca taagagctgt taattctagc 660
 cataacttht aataaatgta ttggttgctt gtgtacatga ctatctgtaa gtaaaatgaa 720
 ggtctcttag aagttaatac agtttaacct taaaatctgt tctaaaatta tttgacattht 780
 ttctcactga ataagaatga gaaggaggaa gcatagtgtg gaaaagtagc gtgcagggtg 840
 gagtgggtact ggattgtaat tatgtaagtht aaggaaataa catgctttgc ctattccttg 900
 tcaccttht tttctgcctt atagacaagg gaaaaaaaag attgaatana agagttthta 960
 ttttaaaaa 969

<210> 63
 <211> 2081
 <212> DNA
 <213> Homo sapien

<400> 63
 ggggtgagcag ataggggaaa gaagagaaca cgaaaagaaa aatagaacat atctgacaga 60
 ttgtaaacca agttaccagc attgtgtgga atcgtagttc caagtcatat aataccaaac 120
 attcaagaaa actagtgcc tgttaactgg gggaaagcaga tttatatcct ttttgaagaa 180
 agtcacctta ttctgtgtat cagaaatctc tttcaaataa cctttcagaa ctatagagca 240
 tcatacaaag atgaccaagc atacaagaaa acaaggcaca gtgagggaaa cacagcagga 300

44

| | |
|--|------|
| tacagcagaa aatagctctt atatatttca gatgacattg ttatttatatg tagattatag | 360 |
| aacagtttca ctttctatgt ttttaagaaat aagcttgaaa atatcagtag gaaacaggaa | 420 |
| gttatatagg aattgacata gcagagttga atgaaatact tcaagatgtg aaaaaatacg | 480 |
| atgacccaaaa ctaattaatg cagtaaacat ggtggaagta gattttccat aaccgcagag | 540 |
| agaatttgtg aaccaaagat agaagaaatt atctagaatt caatgcaaatt attaaaaggc | 600 |
| agaaaatatc aagaaagggt taaagaaaag gaggatcatg tttagaagtt ctagcttatg | 660 |
| cttaattggc attgtagaag acggaaacag agaatcgggt agaggcaata tttgatgaga | 720 |
| taatgacaca attttctagg attgatgaaa gccaccatc cttagattta agaaattcaa | 780 |
| atcccatgca gagtaaatgt gtaaaaaaaaa aaaaaaaaaa aaactctata tagacatagt | 840 |
| gaaactgtgc agaaaacaaa gacaagaaag acgtaaagaa actatcctca aaagagtgtc | 900 |
| atttcctaat aacaaatgaa atctggagtc aatgaaatga tatcctaagt gtgctgttta | 960 |
| ggaaggaggat agaatgcaac ctagagttct ataccagag aaaatttcga cagggaagc | 1020 |
| ataacacctg tatatcagtg ttacaatcaa accatctact gctttggaag atagggccaa | 1080 |
| gaccaagtca ggctgccatg gtaagcagtc ttcacgaat gaccagatga tctctacagt | 1140 |
| cttcatgtgg acgatgatct atgtctaatt tatcagtaaa tatcagcatc tagtaggggtg | 1200 |
| ctttcgataa atttggttga atgaaagaaa aaaggcaaat taaggatattt attcagtaaa | 1260 |
| caaaaaggga tagtgaagaa ccaagctcca ggacttttga agatagactt tcttgtctgt | 1320 |
| ggtgcactga gaatcagctc tatcttagta ttcattttacg tgtggagatc tagaacagag | 1380 |
| ttacattctt tctggtgaca agctgagact acctggctgc tcaatatcca gctcctggga | 1440 |
| gacaagggcc tcaaaactgt tgggattggg ggggtgggca ggtagggaaat cttactacca | 1500 |
| tttggttagga tttttaccat caaacctagg agtagcttgg aactgggtaa gatcttagaa | 1560 |
| tttagagaag ccatagcctg aacaccacac acttctcact gaggaagata ggggtgataa | 1620 |
| gatcatggag gaactggaga ggaatatgaa gacagtgcct acaaccacca catacacagt | 1680 |
| aaacattctc tcttctcagt ggttgaagtt gttcttgatt acagctctct tatctgttct | 1740 |
| ccctttgatt tgctgactga tggatgcagt ccatttaaga aatgagcata gcatattgga | 1800 |
| aagagcacta gacttagatt ctagtcccag ttctgtcact taacagctat gtgttgaggt | 1860 |
| gaggtgaggg aagagtactt ggatctctgg atctctttta actttaatat tctcctctat | 1920 |
| aagggtagaa gatgaactag aggatcacia cagtaccatt tagttttccg ttcagtggca | 1980 |
| attttgacat acttttttga gcatacataa tctcagaatc ctatgcctca aaactttgat | 2040 |
| tcattggaca gatttcaaga aaaatataat ttaacaaatg t | 2081 |

45

<210> 64
<211> 148
<212> DNA
<213> Homo sapien

<400> 64
caatggtaga atgggatgaa atgcagacag atggcaagat gttgtgaggt caaatggctg 60
tggcacttag ttgctacaga aacaacagtt ataaaaatta tgattattac ctagattctt 120
ttgatgatga tgaccccgaga cagagaac 148

<210> 65
<211> 362
<212> DNA
<213> Homo sapien

<400> 65
acaacattta ttaaggtctt gccaggcttc ctgctaccag caggtatcag tcaactgggct 60
gtgctttttc cacacttcag taggttgggt tcatttcatt ttcattgaaa ccctatatgg 120
gaaaatggtt tgaaaatgaa atgagcccag cccattgaac ctattacaat tctcccttt 180
ttgtggatga aggaactgag gccacaaatg gttaagcaac ttgcctaggg cccacaaact 240
gatccttggg agatctgaag tcccagagct tctgctatct ggaccacact tcagcactgt 300
ctattatttt atgctctcca gataccaggg acatttaaaa ttgttttcag cccaagcaag 360
gg 362

<210> 66
<211> 1446
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (218)..(345)
<223> n=a, c, g or t

<220>
<221> misc_feature
<222> (1073)..(1276)
<223> n=a, c, g or t

<400> 66
tatctgaatc catcccccta ctctcttctg ctatgtgccg ggcactctca caagtgtcag 60
tggttgggca cagtcattca ccttttcacc tctagattcc tccttcattc ccatcactga 120
cccatcttgg ggtgggggtgc agtcaggaac atcctgtttg acccttgaga ttcaaagaca 180
gatgactggg aatacagaac accatcccaa aggtatgnnn nnnnnnnnnn nnnnnnnnnn 240
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 300


```

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnagggg tgtcctgagg 360
acaaaatgtg gagaactcag aacttgttgc ctgggtgcata ttggatgctt aataaatgtg 420
gccattatta cctcctctct gagaaagtct ggctaataaa ccatgttgcc tttcaagata 480
gtttttttct accacgactt taaagactgt cttggaactt acaggcaaaa attgtaatat 540
ttgtgctcct tgaataatct cttgcctttt agagggtgtg gcatttagga ggcaattttt 600
tctaagactg gttgcgggca cctgaaagtg caagggtgga ggagaaggaa agaagcctca 660
gttaatgttt taaaatgatg tcctgccaca aacattgacc tctgttcata gcctaaggaa 720
ttctgcaccg ctgcaccctg cagacttgtc agatcttctc tccagctcat attcctccct 780
ttttacactc agtcaccttg tccttgattc ttatctccaa atacattata ttttcttgtt 840
tagggccatt gtacttcctt ggaactcttg tgaaagatca gttgcctata cttaacgtgg 900
gtctattttc tgggatttgt attctgggtc attgacctat ttgtctttta ctaataccac 960
aacatcttga ttactgtaac ttacagtaa gtcttgaagt gggatattgt cattcctcaa 1020
acaatattcc acttcttcaa cagtgtgttg actattctag gtcttttgtc tcnnnnnnnn 1080
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1140
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1200
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 1260
nnnnnnnnnn nnnnnntaaa attatctggc agaaactgga aaaaacttca cagtgtgaac 1320
ctttacacac tattttgtct tccaagtgga agaagcatac taacaatgcc accaatccac 1380
catctcagga aaaaaattaa taaactgtct tttctgtaat ccccttttca atgggggtga 1440
ggtggtt 1446

```

```

<210> 67
<211> 642
<212> DNA
<213> Homo sapien

```

```

<400> 67
cggacgcgtg ggataacatc caattctgag atgaaacagt ctaaaatagg agtaaagcat 60
gcactacttg tgaagtgtga aatggagact ctggactttg ggtatttttg taaaactttt 120
gatatttctg tatacattta aaaaatcaat tgccactaca gtgagttcct taagaataat 180
ctagttatat tttttgaaat cacatataat tagactttat aatatatata ctttttcata 240
tataattaga tctttctttg taatttcata tgtagttcct cataggggtct cagatacaat 300
ggtttttata attgacatat tgaaaaagta tatgaacata atgaaacacc tcatttatTT 360
gataattcac taacgtttta tattcatata ttaggaaagt gaacttagca agcttttttg 420

```

47

aattttaagga tcacatttag acatctcatg ggctgatgaa cacagctggt atctttgtgg 480
 agctttctaa tttaaaaaat gctttgtaga cccattgtc tttaaaccat acaacatgcc 540
 cgtgaagctg atctgggtggg tatgtttatt cttgggtttc agtagggaaa tagcagttca 600
 gagagaggaa gctctctgtc ccagcaccga gactcactcc ta 642

<210> 68
 <211> 325
 <212> DNA
 <213> Homo sapien

<400> 68
 agactgtaga gaggaggtg ggcaagggt ggtgacacta gacaggtgca gtaggtctgg 60
 accatgggtg gccttgcgct acacattaca gagctcaggc tttttttctc cagggtgagag 120
 ggctgggtgcc actgaggcat caagcagagg tttgagatct ccttgggtgac agtgtagagc 180
 agacaggtag atttgggaat ttaagcttag actcacgttg gagactgaga tagctcatct 240
 gagaggcact cagggcctaa tctcaggcag taatttttagg gatgtagggg aagagatgga 300
 ttctgcacat acttgggagg cttgt 325

<210> 69
 <211> 1117
 <212> DNA
 <213> Homo sapien

<400> 69
 ttacaacccc acgatcactc agtcccaggg tgactccatg tctgagagga gaatgccct 60
 gttgtgaggg ggtgacatga tgtttagat cactactgag gggaggtagg gagatcaggt 120
 ggcacaggtg gggcaacagg aaggcagtga caggcagaaa tacattctga gctcatccct 180
 tctgttttct tgtggcagga attttgagtc tagaggaggc aagcgggaag atgtacacca 240
 ggggagggga aagctgccag tcttccttgc ccacagtctg ctttgatttg attcagtcat 300
 ttgatgttaa agcagaattt tgggttctag cttcctacag agaaaactcc tgtttcctga 360
 agtgatcaaa tgtgggtaaa atctggtgaa ggggggttcc tgattctcag aggactatac 420
 cttttgaact cacagaacag tcacacgcag accacagaat gtcccaagac acccgatcct 480
 gtcctttatt cctggctcag cttgggagga ggaaggggt ccaggccagg gcagctggcc 540
 aagctggcct tcctctaggg caccgcacgc ccttccccc tcgcccacgt ctacacagtc 600
 accatcaciaa ggctcaagtg ccagtcact ggcttccaaa gaaggccaca gaacgtattt 660
 tcttcttgcc tctgaatgtc agcttccccc tgggtgcct ctgagtggt ctgccttctc 720
 aggtctttct tggcatgctg agggcctgga gatgcacagg ggtgttcaa tggcacctgc 780

48

| | |
|--|------|
| cacctgaatt gcctcatagt ctcctgagaa acctcagagg ggaggggtggt gcaccagggc | 840 |
| tcagagagaa agaattctgtc ctgactttct gaccctcct ttaggtaggg agagaagatt | 900 |
| ggcagctggg cttgggtgaa aggcattgtc gagaactctg agccagttgg ctgcagtttg | 960 |
| cgaggggagg ggaatagttc agagatgagg gactggctga cacaatgcag aagaagctgg | 1020 |
| gttaaccagc ccagtgactc tcgggactgg tgtagaagca tgaagtggg gagaggcgca | 1080 |
| tattacctat gggcagaaga ggtgttctctg ttgagga | 1117 |

<210> 70

<211> 1753

<212> DNA

<213> Homo sapien

<400> 70

| | |
|--|------|
| gatcctgctt aaacaaggtc tcatggttct caccctaggg aattttgctc tccaagggac | 60 |
| atgagacaat gtctggaggc atttttggtt attacactgg gatgggggtg tgctgctggc | 120 |
| atctcaatag gtagagggtca ggaacactgc tcaagattct gcaaggcaca ggacagcacc | 180 |
| cccacaataa tgaattattt gattcaagat gtcaataatg ccaagggtga gaaaccctag | 240 |
| ggtagctgta gtgttttctt ttaaccttag gggatttgag agccactgga aaccaaagag | 300 |
| ccaaactgaa cctatgtggc tgcaactttt ttagtttctt attaaagaaa aagttctctg | 360 |
| gtttttcaac actaagttcc cttatttcga aaacaaccga aagagctcaa ttttgcaaaa | 420 |
| atggcacaac cacaacagag tgcagtgtga taacaatata gagtcaacta agattaataa | 480 |
| ttggtctact aagacttttag atataatacg ttttattcat ctagtggtag ttggtgctca | 540 |
| cagggcactg tacaacaaat tttaaaaata caatatgaaa aaaatcaaac agccagagat | 600 |
| caattccatt agagggcgat ggcccacaga aaaaccaaca aaggcccaca gaaaaacca | 660 |
| ccctgtctcc ttaagaacat ttctttgttt ggctttaaga aacacaaaaa gaaaatacaa | 720 |
| aagaaaaggt ctttgagaag ataggtaaaa cttcacatag cgtcttcac atatcacaaa | 780 |
| atgctttcct tatttaataa taagcacata aaataatcca aatatctgtg agcctctatt | 840 |
| tgggtccaat gttttagggtc atagctttca ctactccct cctctctatc ctcaaattat | 900 |
| ttgcatcagt aaaaaaagtt aattaagggtg taacagtctc ctttgctcca caatgccact | 960 |
| acaaagccta ttttcataa agaaccagag gccaaagaaa cgtagggtac taagtacttg | 1020 |
| aattgaaaat acaactaaaa tttcctgagt catttaacat caaattttgg ctttgtaatc | 1080 |
| tatttgatt caattctcag tcattactga aggtcatttg acatcaacca tagagagcag | 1140 |
| tgagatttaa agttgcatgg aagtaccgct gaaaacttct aacaacagaa agatgagttt | 1200 |
| ggagtgttta attttactac atagtgcacat atattcagaa agttgtcata caaggaatga | 1260 |

49

tgaggaggtta caaatccaac ttggaaaccc ttaatgctca ctatcaatth gttgaaagtc 1320
 aatththththt thththththt agacggatct ctgctcgccc aggctggagt gtagcagcgc 1380
 tatctcgggt cactgcaagc tccgcctccc gggttcaagc cattctcctg cctcagcctc 1440
 cctagtagct gggactacag gcgcccgcga ccacgcctgg ctaatththt gtatththtag 1500
 tagacgggac ttcaccgcgt tagccaggat ggtctcgatc tcctgacctc gtgatccgcc 1560
 cgcctcagcc tcccaaagtg ctgggattac aggcattgagc caccacaccc ggccacaagc 1620
 tgtthcttaaa ctaagcttct gaagcaggaa ctctaaagt gtcaataggc agththtactt 1680
 acagtagtag ttatatacta atgctgcttg cagaatththt thtcccttha gaggccacag 1740
 aagtactacc aca 1753

<210> 71
 <211> 198
 <212> DNA
 <213> Homo sapien

<400> 71
 gatcaattgc attgatatag atagaaaatg atttgatga tgttatgttc atttatctc 60
 tgaagthtga ggtccttata cacaagcttc cttagacctt tcccaccacc tccaattht 120
 gtaatgtaat atththththt thththththt tggthththt aatgtctgta tataagaaat 180
 taggaggctg ggtgcagt 198

<210> 72
 <211> 7482
 <212> DNA
 <213> Homo sapien

<400> 72
 gcctgccgag ttccgagcga ccgatggaga tggcggctgc ggctgagtga cggacgggtg 60
 agggccagag cccgggctg aaggggggga caaacctggg tgcccgagg agccggcgag 120
 ggtgtcttac aagtatcaag aacttactat atgtgggtga ataaacaatc aaggtaaaga 180
 gcatcaagta aaaacttctg cttgttgata agtacttcag acattcccc agtggctgaa 240
 gtggcatatg aattatgaag ttggatcatt tggaatgaat gtaagagaat tgccaagggc 300
 tctctctact ccagagagga aacctcatcc agggccatga agccacttcc tcaccatctg 360
 tgtgctgctt aagctaagc tgcgggaacc atggttcctt gggaggaatc aagctgactc 420
 ttggcatgag attcctgcct tcctagggtt gagagcggca ctgccatggc ttctctggac 480
 gaccagggg aagtgaggga gggcttcctc tgccctctgt gcctgaagga tctgcagtct 540
 ttctatcagc ttcactcaca ttacaggga gaacactcag ggaagaccg tgatgtcaaa 600
 gggcaaattht aaagtcttgt ccagaaggct aaaaaagcaa aggacagggt gttgaaacga 660

| | |
|--|------|
| gaaggggatg atcgagcaga gtcagggacc caaggatatg agtctttcag ctatggaggg | 720 |
| gttgatcctt acatgtggga accccaggag cttgggtgctg tgagaagcca tctttccgac | 780 |
| ttcaaaaaac accgagctgc tagaattgac cactatgttg tggaggtcaa taaactaata | 840 |
| atcagggttag agaagctcac tgcatttgac agaacaata ctgagtctgc aaagattcga | 900 |
| gcaatagaaa agtctgtggt gccttgggtc aacgaccagg atgtcccttt ctgtccagac | 960 |
| tgtgggaata agttcagcat ccggaaccgc cgccaccact gccgcctctg cgggtctatt | 1020 |
| atgtgcaaga agtgtatgga gctcatcagc cttcccttgg caaacaagct caccagtgcc | 1080 |
| agcaaggagt ccctgagcac ccacaccagc ccagccagt caccacaacag tgtccatggc | 1140 |
| tcccgccgag gcagcatcag cagcatgagc agtgtcagct cggtcctgga tgagaaggac | 1200 |
| gatgaccgga tccgctgctg tacacactgc aaggacacgc tgctcaagag agagcagcag | 1260 |
| attgatgaga aggagcacac acctgacatc gtgaagctct acgagaaatt acgactttgc | 1320 |
| atggagaaaag ttgaccagaa agctccagaa tacatcagga tggcagcatc attaaatgct | 1380 |
| ggggagacaa cctacagtct ggaacatgcc agtgacctc gagtggaagt gcagaaagt | 1440 |
| tatgagctga tagacgcttt aagtaagaag atcttaacct tgggcttgaa ccaggaccct | 1500 |
| ccaccacatc caagcaattt ggggctgcag agaatgatca gatactcagc tacacttttt | 1560 |
| gtgcaggaaa agttgcttgg tttgatgtca ctgccacca aagaacagtt tgaggaactg | 1620 |
| aaaaagaaaa ggaaggagga aatggagagg aagagggccg tggagagaca agctgccctg | 1680 |
| gagtccagc gaaggcttga ggaaaggcag agtggcctgg cttctcgagc ggccaacggg | 1740 |
| gagggtggcat ctctccgagc gggccctgcc cccttgagaa aggctgaggg ctggctccca | 1800 |
| ctgtcaggag gtcaggggca gagtgaggac tcagaccgc tcctccagca gatccacaac | 1860 |
| atcacatcat tcatcaggca ggccaaggcc gggggccgca tggatgaagt gcgcactctg | 1920 |
| caggagaacc tgccgcagct gcaggacgag tatgaccagc agcagacaga gaaggccatc | 1980 |
| gagctgtccc ggaggcaggc tgaggaggag gacctgcagc gggaacagct gcagatgttg | 2040 |
| cgtgaacggg agttggaacg agaaaggag cagtttcggg tggcatccct gcacacacgg | 2100 |
| actcgggtccc tggacttcag agaaatcggc ctttttcagc tggagcccag cagagagcct | 2160 |
| cgcaccacc ttgcttatgc tttggatcta ggctcttccc cagttccaag cagcacagct | 2220 |
| cccaagaccc cttcacttag ctcaactcaa cccaccagag tgtggtctgg gccccagcc | 2280 |
| gttggccagg agcgcttacc ccagagcagc atgccacagc aacatgaggg gccctcctta | 2340 |
| aaccctcttg atgaggaaga cctctccagc cccatggaag aggccactac tggctctcct | 2400 |
| gctgcagggg tttccttaga cccttcagcc cgcacctga aagagtacaa tcctttcag | 2460 |

| | |
|---|------|
| gaagaggacg aggaggagga agcagtggca gggaatccat tcattcagcc agacagccca | 2520 |
| gctcctaacc ccttcagtga ggaagacgaa catccccagc agaggctctc aagccctctg | 2580 |
| gttcctggta acccctttga ggaaccacc tgtatcaacc cctttgagat ggacagtgc | 2640 |
| agtgggccag aggctgagga gcccatagag gaagagctcc tcctgcagca gatcgataac | 2700 |
| atcaaggcat acatctttga tgccaagcag tgcggccgcc tggatgaggt agagggtgctg | 2760 |
| acagagaatc tgcgggagct gaagcacacc ctggccaagc agaagggggg cactgactga | 2820 |
| ccagcagtgg agagggcacc tttgggcca ggggtctggc aggagccagt ggagcaggac | 2880 |
| agagggcagg caggatggat ggggaagggtg gcagggtgag aactcagatg cacacagggtg | 2940 |
| aggggcagga atctgctgtt ttgtgttgcg cactttgagg tatttccact acagttgaat | 3000 |
| aataaaatag aaactagaac agggagaatc agcattcagt tgctgctttt cctgtttatt | 3060 |
| attactatct tttgtaatcg gaggtttacc ccttttgaag ggactttaca tttttactac | 3120 |
| cgagatataa ctaaatgcag ctctgttggg ccagggcag aaatggctgc tgtgtacctc | 3180 |
| ttgggtccat ttgctactgc ctagtcttgg ttccttatgc agtattatag ggcagccttt | 3240 |
| ttagagccct tcctttagcc aagacagaga agatagattc cactgagctc tattctgctc | 3300 |
| tgacagaagt ccatccctag taggctgtga gttccatttc acctggggcc gcctctcccc | 3360 |
| tgctctgcac ttctgtctg tacaatagaa gggggaggtg ctgctatgaa ggggagagtt | 3420 |
| tagaccagg agagcccagc acctctcttt aagggtgggt gatgggaata tttcaccagg | 3480 |
| gtctattttc tcagtttaag ttcttttttg tctctttcag gaagttaagc tcccagtgc | 3540 |
| gggtatcaat gtgaatctgg tcctgagctt tttagaaaat aagagtgggt gccgggcgcg | 3600 |
| ctggctcacg cctgtaatcg caacactttg ggaggccgag gcaggcagat cacaaggcca | 3660 |
| ggagatcgag accaccctgg ctaacacagt gaaatccgt ctctactaaa aatacaaaaa | 3720 |
| attagccagg cgtgggtgtg ggtgcctgtg gtcccgcta ctttggaggc tgaggcagga | 3780 |
| gaatgctgtg aaccggggag gcagagcttg cagtgggccg agatcgacc actgcactcc | 3840 |
| agcctgggca gtgagagtga gactccgtct caaaaaaaaa taagagtgggt taactttttg | 3900 |
| gttgatatta gtattatttg tgagaagagt tccccaccc ttcttctgat agaacaattg | 3960 |
| tctgtcactg gagaaatctc cctccagagc tttggcaaag ttactctgat ctgggtctgt | 4020 |
| ttaaaaggcc gggctctctaa tttaggaatc ggtgatattg aagcttttgc agaacatcac | 4080 |
| cagaagaagg gaagcttccc agagtcagaa gctaaattaa aataatttcc aaggctattt | 4140 |
| gatcagcctt ctcccttttt gggtcattgt gtccctgactt ggggcactga tgagattttt | 4200 |
| tatttttttt gagacggagt ttcactttgt tgcccaggct ggagtgtagt ggcattgatct | 4260 |
| cggctcactg caacctccac ctccctgggt caagtgatat ttgtgcctca gccccctgaa | 4320 |

| | | | | | | |
|------------|-------------|-------------|------------|-------------|-------------|------|
| tagctgggat | tacaggcgtg | ggccaccacg | cctggttaat | ctctgtat | ttggtagagg | 4380 |
| tagagtttca | ccattttggc | caggctgggt | tcgaactcct | ggcctcaagt | gatctgccac | 4440 |
| cttggcctcc | caaagtgctg | ggaatacagg | cgtgagccac | tgtgcccgcc | ctgatgagat | 4500 |
| attttattac | caatgttagt | attgagaaac | tgaaatgttt | gaagaagcac | aacccaggat | 4560 |
| cgtgctggta | gcaccacagt | acttaaactg | ttggtcaatt | aaggccagaa | agggaaattg | 4620 |
| ttaatttagc | tctgggtgctt | tggtttacag | gaacataact | cttaactgac | atctgacatc | 4680 |
| atgatagcca | tatgtgctca | gctctgggta | gagtttctgc | agttactcat | ctgaactaat | 4740 |
| gaacaataac | tgaccactag | tcactttatg | ccgtgtaact | agctctaggc | catactttca | 4800 |
| ctgggtactg | gtgtgaaagc | tgaaattcat | tttggtacat | tctgggtgaag | acccctcttg | 4860 |
| ataatgggaa | tgttttaact | ctcttgatga | aaaaataatc | tgtatttggt | ttgatgttca | 4920 |
| catttctgta | gcacatttct | tatccttttg | gttgaatgaa | aagatcttgt | ataggggtgt | 4980 |
| ggagatgggg | agtgggtaga | agtgtgtgaa | ggacgctttg | catttgggat | ctgttcacaa | 5040 |
| acagccatat | gagtgtgtta | atgaatgtca | gccagttacc | aaccctgctg | gttggttatgg | 5100 |
| gttgttttga | gaagttggca | accaggcatc | taagatgttg | cctggtacag | gcctcttttc | 5160 |
| ttccctgagg | cccatgacat | ttctctgtta | ctctagaggg | tttctcagat | ggccagtagg | 5220 |
| ctctccctg | cttaggggtct | catttctctg | aaaagaggat | gaactgaaaa | gcaggtagtt | 5280 |
| ccagaagcta | attgctgctt | tccatcataa | ttatttttct | tgtgagaaca | tttctctttt | 5340 |
| aattagctag | tgattttgat | taagactaat | tcactaaaca | tacccttccc | tcaaatcacc | 5400 |
| tcaggtagca | atctgtaagt | aactaaaagc | attgaaaaac | acacaagaaa | cattttttaa | 5460 |
| aactattttt | aaaggcctgg | ccgggtgcag | tggctcactc | ctgtaatccc | agcactttgg | 5520 |
| gaggccgaga | cgggtggatc | acctgaggtc | aggaatttga | gactagcatg | gccaacatgg | 5580 |
| tgaaaccccg | tgtctactaa | aaatacaaaa | attagctggg | cgtggtggca | gacgcctgta | 5640 |
| atcccagcta | ctcaggagg | ctgaggcacg | agaatcgctt | gaacctggga | ggcggagggt | 5700 |
| ctcctgcctc | agcctcctga | gtagctggga | ttaccggcac | ctgccaccac | accgggctaa | 5760 |
| tttttgattt | tttagtagag | atgggggtcc | accatgttgg | ccaggctggg | ctcaaactct | 5820 |
| tgacctcaag | tgatctgccc | acctcggcct | cccaaagtgc | tgggattaca | ggcgtgagcc | 5880 |
| actgtgcctg | gccacgaagt | tcagaccgta | gagtttttca | taatgcaatt | gaaaccttat | 5940 |
| attcttatgt | ttcgacagg | ctgggtactt | aacttaaact | tttgaaaaaa | aaattgaatt | 6000 |
| caactctcag | aaagcttatg | gccttttgca | gaattataag | tttacaataa | cctgccatgc | 6060 |
| aacttagtga | tagatcagat | tagaaagggtg | cagcatgctt | cttgtttttag | cacgcctgaa | 6120 |

53

ataaaaactaa aagaaacaag agtccctagg ttttggtagg ctaaatttgt gtcttggtt 6180
 ttgattctac tgtctccctt gtttttgact ctttttagtaa ccaagcaacc tctagcctca 6240
 gtcttcattc tcagggatga actccttaga gtgaagtctt cactggtagt tcccagcatt 6300
 ttgaaatctg agaattgctt ttaaagtaaa aaatgttggc tggcctccac acaatagtag 6360
 ttttgaagtg tttgggaact gcatcatgat gaaaacaatc agatttgtat ttttaatcat 6420
 tatagcagaa gctacaaaca ttttaagaat acaagtacac atgtggacaa aacattcttt 6480
 attctataga caatgcttgg tatgcatatg ggtttgataa cccatttttg ataattctag 6540
 gaccctgcc aatcatagaat tagagtcttc tttttttttt ttttttaatt tgagacaggg 6600
 tcttgttctg ttgccagggc taaagcgagc tggcgagatc atagctcatt gcagcctcaa 6660
 actcctgggc tcaaccaatc caccgcctc agcctcctga atagctagga ctactgatct 6720
 gcaccaccac acgtagctga attataatca ttcttgggtc tacttgggtg aaagtgaagt 6780
 gtagagacct taaataatth ggtcatttga gccaaatgcc aacttctctc ttctgacttg 6840
 ccctcttttt ccctaagaaa tttggcagtg tagaccgaac tgtctgctta attctgttcc 6900
 tggatctaca aagttaaagc gaggcccaag gattactcaa gcaagtaaca tcctttgagt 6960
 tatctgaatc tgtgtctgtt ttctgactga ataagatgca gagacctggc taaatattca 7020
 agtaagccaa gaatagaaca gaaagatcat ggtattctct ttaaactata gtcaagattg 7080
 aaaggaagat gaatctatgc cttcctcttg cttgggttagc cttttcctgt tttctactta 7140
 taaaaatcaa agggtaagac agacaatagg ctagatttaa gataggccca tacaaggatc 7200
 tcagcaagga tcagaatgtg ttatttctgt acccattgta gtgatgtcct tgctgtcaaa 7260
 gctacagaag atcaagaata agtcttcaaa tgcaagtctg tactactgaa tgaacaactt 7320
 taaactaaga tgagggactt attttctttt tgacttgagc aaaaggctgt caccttggtg 7380
 ttctgtttct gttgtaccct accaagtaag gaggataata atggcataaa catatctcct 7440
 tgacactagt taatcatgtc caagctcttg gatgaatccc ac 7482

<210> 73

<211> 2512

<212> DNA

<213> Homo sapien

<400> 73

cggcacgagg cagccctgct gaagccccag agctcagccc agtgctgggc acatagtagg 60
 cgctcaggaa atgattgctg aatgtggaat gaaagtcccc agcaggcaaa aggctgtgac 120
 ttctgttagt ggggtggtgg tcctctgggg cgagtgtatg gagaacttca aggcctctc 180
 ctggctgccc aggaagtaga gggcgccaaa ggcgacgtgc cggccgcccag ctccaggccg 240

54

| | |
|---|------|
| agccccgagc gctgcagga acaggccccct tcacccggcg cgggacgcag agctgcgaga | 300 |
| gaatcttggt cagcgggac tcaacgccag ggcggccgct agaggttggt ctctgtctcg | 360 |
| gcctcacccg ccgggagacc acagagctgc ttccccagcc gcccgctgcc agaaattgga | 420 |
| aaaaaaaaa aatccagctg gggctctagga actcggcttc tggcacctct gaattctccg | 480 |
| agactgtctc ctccctcccc gcctgtaatg aaccctgtga agggagacag gccaggaagt | 540 |
| cccagaaata ttattctctg tgactctcac aaaatggaaa agggctctca tttttgtttc | 600 |
| tttaaggaac ttgtgttctg cgtctgtgtc tacactgcct cctctacca accaaattgt | 660 |
| ctagcccccc tccagttacg ctagaactct gctttatctt caaggaagaa agggagtggg | 720 |
| gagaagttac ctctaaaccc tccagcatgg ccatcaattt tctgaataat ttggagggtca | 780 |
| acatgctttc ggaaaagtgt ttggaaaacg tttgggggtg aaaagaaatg ttgcctaact | 840 |
| ggcctttact cacggagtga ccatcatctc agctagttag tgtatgagcg gcaccagcac | 900 |
| cccggccatg ggtgtctgtg ctgagcctct gaaagttgac ttatcctttg gggagccctc | 960 |
| ggaaagacac agctggtatc tatgggaatg gctttggggg gtcctttggc accatacaga | 1020 |
| taattttgtt tttttaattg gaatccatgg cctgggttct tgggatgggg gtagggggaa | 1080 |
| gccacaaagc ccctggaaat gtatgcaaaa catatgggtg tgcaggtgca tcatcctaag | 1140 |
| gctatggcct ctacagtttt tccaaagagt ctaagactat aaaagggttc catctccttt | 1200 |
| tctcagccag ggagccacag agaccaggc cagcagtagg ctgggagcca gcacaccggc | 1260 |
| ccccaccagg tgcccggtag gggatgggta ataacactct tgaatggtct ggaaactgct | 1320 |
| ggccctcag tcctgccgtc ctcccagagg gaaatcctaa tagaagaagt gacagccagt | 1380 |
| agctgccatc tcagtatttt gaatgaaaat ataacatgat aaagagcaaa tgagggaggc | 1440 |
| tggagggggc tccccatgca aatactggaa tatgaatggt gccagaaat aagatacttc | 1500 |
| ctgccctcca gctgttaaag agtcccca ttcctccctt ttggtgctag gaaggggcca | 1560 |
| gagggcactg ccccagacc accaccacc tctggagagc tggggccgag tgggactgcc | 1620 |
| tgcagcagag ttaacccttc agaaaattca aaacctatac tgcctagtat ttccctgcc | 1680 |
| cgtttatttt ttaatcagat ttttaaaata tccagcagca ctttggtgcc ttttcttggg | 1740 |
| gttggggagt ggggttatct atgtgtgtat ctatgggaag gtgggtgcct tcagtgcact | 1800 |
| aaacccctc ttgggcctgc agctggacct cactcctgta cctgaacact ctgcctgagc | 1860 |
| ctgagctaatt cttagccatc cattgtgtct caaaactggt ttggagggtga ataactgtga | 1920 |
| ggcaggatgg agaacccttt gctctcccat ccagaagggc acctaaccag gccctggag | 1980 |
| cagacaaaag gagcaggaag tcaatcactt cgatcccagt tctctgaagc ccaagaagaa | 2040 |
| aacggatttt cttctgtttt ggttcggagg cctagtagag aatttggatt ccaccaagtt | 2100 |

55

```

ctcttttttc aaaaaaagta aacgggtccag agcagacaaa aactgtggaa acttgaggcc 2160
tgggtagtag tggttttggt tgattttgag gcttttaaga gataaggaga cggtaggtgga 2220
gctccgccac gccgcgtggc tctcacttcc acgccgagga agaccgaccg gcccttaggt 2280
ttatcctctg ggagcccccg gcccaatcag ccccttcag gccttgggag tcaccccgag 2340
tcttatcttc tgtccccctc cagcaagtac ccggcaataa acgggcaggg ctgggcaaag 2400
gtctggccgc ggctggtcgg cctctgctct tcaaacctcg agctcactgt tggctgctct 2460
gggccccagt gtgggcagga gctgcgcttt tttccctgct tgcctttctc ca 2512

```

```

<210> 74
<211> 954
<212> DNA
<213> Homo sapien

```

```

<400> 74
gcacgagctc eggccatctg ggcgctgctc ctccgcgcct ctggctcccc gcggcgcgcc 60
agggccatct cctcaccagc cgtcttgga tccgacggcg ccgtgggtgg gggcaggtga 120
gacggccgag ccccgcccgg cgatggcagg aagcgtcgc gcaatcccc tcccgaccc 180
tgtcctcgcc cgccactcac cggcgccagg ggaacgcagg agcgtgagcc ccttcgcgcc 240
cccagcgccg tcggcgctgc tgcccagac acagacactg cctcgagagg cctcacagag 300
gcggggggcag aaggcggcga ccagagccg ccacatcccc cgccttgggc gccgtcacag 360
tcccagacg ccctggactc ctgcagtcta cgaagacgcg cgggggacgg cgtggttccg 420
agagagggcg ccaaaggcga cgtgccggcc gccagctoca ggccgagccc cgagcgccctg 480
caggaacagg ccccttcacc cggcgcgga cgcagagctg cgagagaatc ttgttcagcg 540
cggactcaac gccagggcgc cgcctagagg ttggtctctg tctcggcctc acccgccggg 600
agaccacaga gctgcttccc cagccgccc cgcagaaa ttggaaaaaa aaaaaaaaaa 660
aaaaaaaaat ccccgggggg cttagagaaa ccgcccgtgt ctgggcccct ttagatttct 720
ccagaaaatg gtttctcct ttccaggtg ttataggggg acccccggtg tgcaaggggg 780
aataaccggc cgaggaagtt ccccgagaa atttttcttc ctgggggacc tcctcacaat 840
ggggccacga gggctcccaa cttgtggtgc ctttacaggc gcccttggt gttcccggcg 900
cctgtgtgtt caccgagggg ctctcttct cccactcat tgtgtttgg cccc 954

```

```

<210> 75
<211> 949
<212> DNA
<213> Homo sapien

```

```

<400> 75

```

56

```

atcacttatg atggcaaaaa aaacacaatt acctttgcac caacctaaaca ttaaactcag      60
gcgctttaa ttttactctg aaatacgttg aaagctagct tagttgatct ctgtcattaa     120
cagaaaccat ctctgttttt gttattttct tggcttgatt taacgttcgt tgctcgcaag     180
aaacagaaac ccaactcaagg tagttgaagg aaagggaggt ttccattcaa gggagttgga     240
gaatcggaag gcaagaatga gcctcaggcc tctcatggga ctggaccctg ccagtggaaa     300
gctgttgtct cacaccctgc cccacctcac aatacatgtc tcttttctgc ctctttcctt     360
gcatgcttgt cattctcact cgctatgtat gccagcttcc cctgcagctc catgcacgct     420
gcgtctgttg ccctttcctg ccttttgcct tccaggctaa aactagatag cttttgggtc     480
ttaattccaa atgcttagga gtgaatcggg tagctcagaa gtccaaccag ttcaaacc     540
tgtgtcaatc tagcacatta attctcaacc ttggctatta ttgatgtca ctgggagctt     600
taatagatac caaagcctgt gtcccattac cagaaattct ggtttttagtg gtctgggatg     660
ttactaaacc taatttggtg ggatatgttt aagttcccag aggattctaa catagagctg     720
tggtccatgg ttaagaagca ctgatctagg agagaagagg gtcactctat aggcgtggaa     780
tgctcagata gtgggggtctt ggcttaggta gattattctc ccaaagatgt cctctagggt     840
gtccccattt ttggctttta cctaccagtc ctctttctta taggggtgtg tggaataacc     900
tatggagctc tactttaagg aaaaaaaaaa aaaaaaaaaa ttggcggcc                949

```

<210> 76

<211> 2563

<212> DNA

<213> Homo sapien

<400> 76

```

ctgcttctgt tacctcaaag tggtaacagg actgatgccg aaatgtcacc aggtcctttc      60
aggtcttcac agtggagaac tcttggccaa aggttttttg ggggaggagg aggaaaccag     120
ctttctgggt aagggttaaca ccagatgggt cccctcattg gtgtcctttt aaaaaatatt     180
tactgtagtc caataagata gcagctgtac aaaatgacta aaatagattg taggatcata     240
tggcgtatat ctgggttcat cttcaaaatc agagactgag ctttgaaact agtgggtttt     300
aatcaaagtt ggctttatag gaggagtata atgtatgcac tactgtttta aaagaattag     360
tgtgagtgtg tttttgtatg aatgagccca ttcattggta gtcttaagct tgttggaat     420
aatgtaccca tgtagactag caaaatagta tgtagatgtg atctcagttg taaatagaaa     480
aatctaattc aataaactct gtatcagccc ccaacatatt atttttcatt atttggggga     540
tatttcagtt ccagagcagc agtatcatgt tttctttgtt ggtgctgtct atagttcatc     600
atggtttacg tgtgttttcg ttatagctgt tgccagattc taaagggtt gatattcaaa     660

```

| | | | | | | |
|-------------|-------------|------------|-------------|-------------|------------|------|
| aaaccacaga | tgcttttcagt | ccagtatatc | ctagaaatat | agagctctac | tttgtgcaat | 720 |
| gcactgggga | tacagtggcg | atactgtcct | tgtcttcaag | gagttcggag | tcctagtata | 780 |
| ggagacatac | ataggagaag | ataattttca | cactgcagtg | gttgtagtaa | tagaatggga | 840 |
| gtccaaaggg | gagttccgga | gaggtcaggg | gtgacttcct | ggaggagatg | cccaagcttg | 900 |
| gaggctggat | aggctttgtt | gaaagatggt | acacaagagt | gtgaaacaaa | attgtgtgtg | 960 |
| cagggaagctt | aaaatacaag | gctggggaaa | gaagttggag | aagcaggaaa | gcccaggccc | 1020 |
| tctagtgtct | tacggaacat | cctgtggagg | tgagagctga | cttgtagggtg | gaagcagctc | 1080 |
| tttggaaggtt | tgatttgga | ggtgaactga | gaagaagggtg | gtgatgcaag | ccggccgtgc | 1140 |
| tgaagccagg | atgaattggt | gtgactgggc | ttcagttagg | ctggcatagc | agttgagaga | 1200 |
| gcttagcagt | gggcagcagg | gcactgttg | gggcgggtggt | gagcgggtgga | ttctggctgt | 1260 |
| ttggaggcag | gacagtgggtg | gaattcgatc | attgattgga | cgtgggacag | agaagggaga | 1320 |
| agagaagact | catctaggat | gagtcccagg | tttctggctc | agtcaactgg | ggaaacaaag | 1380 |
| tcacagagct | agggagtagt | tagagaacac | atctgggggt | gtgactcatg | gtcagttttg | 1440 |
| ggctcgtcag | ttttgagatg | cccaaatac | atgcagatct | gtcccacctg | aaaatggaga | 1500 |
| agacacggga | aaggagagga | gtaaaaacta | actccttttc | acaaagtgga | agttaccaga | 1560 |
| atgtgatttc | agaggccccg | ggggatttat | catgtgacta | ctgacccatc | ccacctcttg | 1620 |
| cccctgcctg | ttgcacagtg | ggcaagaatg | tttgtgacct | ttcactacca | ccacctcccc | 1680 |
| cgagcatggt | ccccccagtt | ttcaatatga | accatcctgt | gggtaccctg | tcacaggctg | 1740 |
| gccctgaggt | gagcaatatt | tggactgtga | tgttggttgt | tctccactct | ttctacagga | 1800 |
| cagaacaggg | cctctagagt | gggaaatggc | tttgggaaat | atgccaaagca | gtagcctttg | 1860 |
| ttcttcaact | tgccaagagg | ataattctcc | acacccttcc | tgtactcagt | cctcagttgc | 1920 |
| ctggtgagag | agcagcctcc | tcccgtgtgc | tctgccagct | ggaccagac | tggccatatt | 1980 |
| accagtgaga | ccaaaaagat | ggaggtgggg | aggtagctct | gaggtctggg | aaaccattcc | 2040 |
| agctcctgcc | agttttaact | tgtgtttaat | tcctggcaca | gttgtcctgg | aaatgccttt | 2100 |
| ttctcttgcc | tgggaaccac | tagaagggga | tgttgtctgt | gttggccagg | gccatgcaaa | 2160 |
| ttcaacatct | tgtttctgcc | cttccccgt | gtagctgagg | ctagggtgtg | gcattacca | 2220 |
| gtgcttggtc | ttcagagagc | aaaagcactg | ctcgtcatgt | ctgaaattta | gtgagtgagc | 2280 |
| tcacccacta | ggctggtgtt | tccctgccgt | ggctgcacat | tggaagcacc | ggggcacttt | 2340 |
| gagaactaca | gatgcctggg | tcccagagca | tctaagggtgc | tctagggtgt | gtccaggaca | 2400 |
| cagccctggt | tgaggaccac | tgctatattg | tatggcctct | tttaaaaaag | ttaattttac | 2460 |
| ttggaaatga | tttcaaagct | acagaaaagt | tgcaagaata | aaaactgtac | aaatgaggct | 2520 |

caaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaatatgcg gtc 2563

<210> 77
 <211> 614
 <212> DNA
 <213> Homo sapien

<400> 77
 tgcaggtaaa tcatgaaagc attgttttat tcttttgtca tatttttcga aaacatcaaa 60
 ctttagagac cagtctagtt tatcttgctt tctcagtttt attcccatTA aacttgatct 120
 tcttgcttat atttagtttt atgtcagctg tttttccagt aaatgtatag cattgttcaa 180
 ctaagttcag actcagtcct tattcatact tgcagttaac tactttctgtt atccttgacc 240
 atctcagtcg ggccaaaaac ccaaaagtaa atgaaaaaag acttctgaaa atggaggaaa 300
 ttagtcttca gaaagccctt tccttttttc tggaactgcc actatcaagt cagatcctgc 360
 tgcttttaggt tgtaaatgtt accagctaag tatctttttt tccctataa cctcaagttc 420
 tttctctgct ttaacagggt aggagctatg gagtaggatac atgaggtggg tcgagagtac 480
 cttagaaaac ataatacata aagctaataa ccaaacaggg aggaaagaca acacaaaatc 540
 actaacatca acaatccacg gaagtctggc gttgtagtgt agggacaagc actcctccaa 600
 tgcatacaaga taag 614

<210> 78
 <211> 361
 <212> DNA
 <213> Homo sapien

<400> 78
 aagaagagtc ttttgtgtgg aggctgggga ggaatgaaat ttgaaaatga taaaaagtga 60
 ttgaagactt ctctaaatta aaggtagtat cgttcaactc gcttaaagag tgagcttgtg 120
 aattaattac agttaggaga tcaagaaact gagacattaa ttgatacaaa atcgcatcca 180
 acacgtagat ctagaagaaa atatctttat tttctttcta gcatcaagga atattaaatc 240
 taaagtgtcc ccaggtatat gggttttcaat cctgtgctta aatatctcct gagaaggaga 300
 ttccataaat tccctctata actaaccaat cctcataatc agcaagttct tttctatctg 360
 a 361

<210> 79
 <211> 1321
 <212> DNA
 <213> Homo sapien

<400> 79
 gagaaaatag cataatatag gatggcagaa ataagtctag tagtcacaag cactgtaaat 60

```

gaattcagta cccctagtaa taacaaacac ttttaagatga gtttgaaaga aaatctagtc 120
atatgaagtt tcataaaaca gtgacataga aagattgaaa gacagctgat tgaaaaatat 180
gtaccaggca aatatcaaca ataagcaagt agcaatatta atgccatact gaatggaatt 240
caagggtttta aatagcggttc agaagcacac aggcataattt tataaaagat aaaattcact 300
atgaagacgc catcaatttc tatgtaatat caatgtataa tgtcatctct cagcttcaaa 360
gtatgaagaa aatgcgatat tttctcatac ctgatgaaaa actttatcac actgctttta 420
gaagctgacg tattaagtag ataatataga tagagatgca aaggggattt ggaaaacagc 480
cttgatccaa aataaataga tattaacctt gtattcaata aaaagaaaag cactttcctt 540
ataaatataa atgggatatt tataaaaatt aacgatgact tgtgccctaa agctgtattt 600
tgccagtcac gttttctttt tgttgtaa atattctac tctagctggg cccctttga 660
agacttgtaa ttttttaatt ggttgagaca tgcccaagt tactcctgtc acagaggtag 720
tggtggccc ctgggtctct gtgtttgtct tgtcattgaa gaatcaggca gcctgccagt 780
cccttcctg ctgagggacc tagaggaagc atctctgtt cccagctctc tcttgcctcc 840
ctgtcttttag gaatagagaa aattcagacc acttgatcag cattggctta gccttttggg 900
atccagaaga gagcaggcag atctaccaga cgtccctgcc aaccctggcc cccaaaatca 960
acagctctct ggtggccaag ttcttgccag tatctgctgc tttcccagcc ttgctttagg 1020
gcagcctggt gcagttctcc aggttaactt agcaaatcct cacactggaa atctcctgtc 1080
caccaagatg tccaaaattc cttttaggcc ttaccagata cctgcctagg agaaaactgga 1140
gctgacaatg gagatgatgg gagttatgtt caagttgcca tattccagca ttcgccacaa 1200
ctcctatcag ttttgccagg caacttacta ggtgtctttc atacctgctt ccttcacccc 1260
agtacctcta ttttcataac aagaagagtt catgaaaatt accctaccaa agggaaacttt 1320
g 1321

```

```

<210> 80
<211> 293
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (155)..(193)
<223> n=a, c, g or t

```

```

<400> 80
agctagtgtc agggagtttg gatgccaga gctcccctga gggcagctta gatctgtttg 60
gtgtgcaagg ccctgtacac agttcagaaa agagaatgag ggggtcaatct agttactttc 120

```

60

aactagtact ctctgttggg ctgtttctct gtctnnnnnn nnnnnnnnnn nnnnnnnnnn 180
 nnnnnnnnnn nnnatgaaaa agacttgact aaaagagaac agaagatttc agagagcaca 240
 gaatgtagta ggagagtgat aagttgtgaa actctttatc tttgtctatc cac 293

<210> 81
 <211> 549
 <212> DNA
 <213> Homo sapien

<400> 81
 caggaaagaa aaatttatgg aaatgacata gaattcagaa agtattcccc acctaacagg 60
 tgggagaaat tatgtggcct attttactct aatattatgc aaccaagaaa gtagaaaatt 120
 ggaagatatt ctgaaaagca tgtataaaac acattgcaag gttacataaa atagtgacaa 180
 aaatgagttc aatatatctg gatctgatct gattatttag tttaggagaa tcgctgttgt 240
 caaagatggg gactctggaa atgctaaact caggcttcta gaatgagaag gggaaaatgg 300
 gcaaggtcac gatgggtctag ctgcttttta aaaatatgta ttaagggatt gaatatgttt 360
 cctagcttca ctgtatatta ttagaaaatc tatctgtgaa attggttggt cgagatttcg 420
 taatctgtgc gtaagtggct gttcaggatg cttacatgct gtcttaatac cgaagatgcc 480
 atcctttttt ccagggccct cttcttacat gggaagatag tggatcctac tataaaatgc 540
 ggatatagt 549

<210> 82
 <211> 282
 <212> DNA
 <213> Homo sapien

<400> 82
 aaataaaacc tatgtttttg agagtagcag gaaacatttt gttagaaaac caaatgaata 60
 cagggagata tggctatcag gagctgagac tgatattata attttccaag tactagtatc 120
 tgtgtattga tttttacagt aatttgtact ttcccatcca gcttgtttgt gaccactgag 180
 gcccttgggt aatatcttag aagtcactca aaatggataa agaatgaatg tgagtcaccc 240
 atagcagtct tatgtgagat aatatatggg tgatectcag ta 282

<210> 83
 <211> 71
 <212> DNA
 <213> Homo sapien

<400> 83
 ccaagataat gaacattcat aatcccagta gtttcctagt gtccttttgt aatcttgctt 60
 ctcaaaaacta t 71

<210> 84
<211> 400
<212> DNA
<213> Homo sapien

<400> 84
agcttccacc ctcccttggt cttgtacaaa taggaatata gtccctccaga agcccatagc 60
tcattcagggtc catgatgtat aaaagggtggg catttaatat cttgtcattt tattccttct 120
gtttaaaaat cttcaattaa cacctaccac cagtcaaatg aagtctaaaa gtagccttcc 180
ctcattagga gtcaaacttc acaacttaat ttctgctgtc tctgtgtgcc tgtacttcag 240
aaaaaaaaa caaactgttg ggtttccttg tgagagtgc ttttccttgt ctgcatattg 300
tgcttctgcc tttccttggt tttctcatga ccttatggca tctctgactc caaattgtat 360
gtatccagca caaatctcaa tccttcacac ataatcattg 400

<210> 85
<211> 165
<212> DNA
<213> Homo sapien

<400> 85
ggaagctgat gtcattgaaa agatattgcc ttgctttgaa atcagcaagc gtggtcaaag 60
ttctccagta tatacacaca gaggaataat gtggctctga atcaccatgg gcagcttttt 120
gataaacaga tattcaaaaa aaaagaaaaa aaaccaccg taaaa 165

<210> 86
<211> 406
<212> DNA
<213> Homo sapien

<400> 86
aaagaaattc ttccacagt acattttcac caggcaaaact cagagttggt tgacatgaga 60
aggtttaaaa aaggatagtt gtgaaagacc aggcaaagtt cacaaagtga agaagggaaa 120
atattatttc agactgatat taaatgatgg cctagcctgc tgccatactg tctcacaggg 180
aaaagtttta tttatgtctc ctttgatctt ttctcctttt tgtctctagg tttgatgtac 240
aagctttagc ttgggccttt gatttacagt gttctggaga aagagaccag ggatgacaaa 300
gccaaaccagg ctgttaccta aaaacctgat atattgtcct agatctaggc tgtcttagag 360
ctacacaaag aatttccagc agactgctca ggaagagcct gctggg 406

<210> 87
<211> 416
<212> DNA
<213> Homo sapien

62

<400> 87
 cagtaactct cttgagaact cagacaagga ttagtcctct cagctgggct actcctgatg 60
 ttgatggtga aacaccagcc actcaggcag tgtccatcaa tgtcaaccct aatctggaaa 120
 tgaactgacc aggccaagggt agaagcctgt tggctctgtcc aactagtaac agatccgtaa 180
 aagtgacagc cactttttacc accatcatta tagctctgac cactttccct gacttgattt 240
 ggccaatgtg ttgtgtgagt ggtggatgga agaagctaata acagctatgg ccttcttcaa 300
 taatggtgag gtaccactat catttgcaaa tcagttcaaa aataggaggt cttcccagct 360
 gggcactaag agctaattct cttgggcaat tttttttatt gcattcaacc atagaa 416

<210> 88
 <211> 67
 <212> DNA
 <213> Homo sapien

<400> 88
 gaaagattca ggaatatagg ccaataactg gttcccaact aggcagcttt cttgaataca 60
 gaaccct 67

<210> 89
 <211> 168
 <212> DNA
 <213> Homo sapien

<400> 89
 ggccagtggg ggcaatcact gagcttggcc cggcgccctgt ctctcctttg tcagtttccct 60
 tttgtgtgcc tccagctgct gggctgggag tgggtgcagag aatcaggggc acacagccct 120
 gctctcaggg agattacact tgggggacag aactggggtg cttttcct 168

<210> 90
 <211> 328
 <212> DNA
 <213> Homo sapien

<400> 90
 attccttgat tttgatcact cttgacagga tttatcattt ttttctttaa aaaaaccctt 60
 ttagtattgt ctaccacttt atatttaata tttctttggt ttgattaaat aataataatt 120
 tttattttta tttcaagttt ggtttgtatc cctgctttac tttccaattc tgtgaagtgt 180
 attaccagat cttaaattgt caggatttcc tcttttgcgt aaatggcatg aattttattt 240
 tatgcccatg ttttgatgca tccacacaat ttctagtaat ttcagttaag tataattgat 300
 ttaaaataaa ctgaaccact ttaacatt 328

<210> 91

<211> 2377

<212> DNA

<213> Homo sapien

<400> 91

```

gttgtttgaa ttttgagcct ccagggttaga ggctccaagc agacaaagaa gaaagtttgt      60
agacacaata ttctaacatt gcacaagttg atgagatatc aagttttgaa tattattcag      120
gaagcacttt ggagaatgtg ggagtcctac tgttttgcac tagtctgtaa tttgttacct      180
ctccccattc gataatatag taactcagaa ttttcatata aaattgaatt ctactcatta      240
gagtactttt taaaagtaca gcagaaagaa aagaagtgga acataagcat tcaagattaa      300
tatttgattc tctatctctt aactgtagat tttattggcc tgtttttttt ttctaaataa      360
ttctttaaaa cttgactgga acatgcctct gcatttctga gtgtagaaat taaatgaagc      420
cactcacagt cctttgattt cccactgaag ataccccaaa ggatgcaagt gcctacagta      480
ttatcaggag gagaacatga aaatattaag acaaaaatcc tcagcagttg ttctccaccc      540
cttctccacc tccaccagca aggagaagtc aatctacact tttttctcat gttctaaagt      600
cttagaatac tgctggattg ttgagcatga gacagagcaa aggttagata cacaaggtag      660
aagttcatag gagcacaatt ccttgattca gggaaagtag cacagaatgc aatttgaaaa      720
ctacaggctt aggactctgt gccaaaaatc tctattttta atgaaaatat ttatataaaa      780
atacatttat tctgcttaac acaaattaaa actgacataa aattttatga taataacatg      840
attcaagaaa tgtgatgtag atttttgaga atgccaaaaa tcagggtttca gaaaagctac      900
agtatgaaat ctgaaaagca acagaaattg aaagtgacct taaagaataa atttcctaatt      960
ttttcccagt ttctatgtt tataactatt ataaagagtt taactttctg gaactgaagg     1020
agaaaacatg aaaatgttta tattgcttct tgttaagcag gtagaccaa ttccgctttc     1080
agtaattcat ccatacagaa tttggatgac tatgtataaa ggaaaactta aatcttaatt     1140
attatcagtt taaatttttt ttgtttggtg aatgacaaag gcaataaaaa taaatttgac     1200
tttagctttt tttctatact ttctctttt tgtaattctt aaattctagt cactagtcatt     1260
tatgagagta ttgactaaat attttcacaa tctaaatatt gtcacaatct aaatttctgt     1320
agtagagaga gccgtagccc tttgtaaagg ctttcgttcg tccaaacaac acttgatcca     1380
accaaagttc caatgtgact gtggatcttt gataggctct tgggtgttgg tgtacaacaaa     1440
tttgatttga actacctata aatgaaaagt cgggggtttat tactttatat gtaatgctga     1500
gaggagactg ttgggaacag ttatgtcaat gagcaaacta gaaattggct tcaaagctctg     1560
atacttttaa aagcatgtgt tttgtgtata tgctcacaaa atgtctgtga agagatttct     1620
ttcagctttt gctcagcttt atggtggggt gtactgtttt atctgggcgg tgagggtatt     1680

```

64

gatttacata aaaactgtag acaaacacag tacgagtctc agcacgtttt gcattttattg 1740
 tactgcccaa attgttttta tgttaaaagt cacgtttcat aatctgcaat atttttgtca 1800
 aagtgcacac tgtactttct tcttaaaaac gtgattaaag agatcacttt gtccggatga 1860
 agcagctggt ggcatatcc gctgaggctt gaacaccacg tgctgattct ctcataatc 1920
 tgaaaaataa cactcgtggc tcagaacaca caagggtaaa gtgtgtggcg cactgctact 1980
 ttttatctca ctttttaacc aaccctttct tcaaccctct tctaaggaaa gcagaagtcc 2040
 cctttgggta ccagctgtgt ggcataacca tggctccttg tccctgacag atgacacctt 2100
 cctttagaat ccaccagcat catccactct gcgtctagct gacaccctca gacagggatc 2160
 agccaccctg gggcgtgcag gtcaaataatg cagatgggga ctttgacagc cacttcagt 2220
 tcacgctggc acatcatttt ttgatccttt ttccttttta aacttctgc accttttcag 2280
 ataaaatcat attttactat gtatagctaa taaatcgtag ttttgttctc ttaatcgcat 2340
 ggtcaggaaa tgagtttttc taggtgtaga gtttaacc 2377

<210> 92
 <211> 182
 <212> DNA
 <213> Homo sapien

<400> 92
 ggacgatgat cagaggtcac agtttgggag ttgccacaag ctaggcttac ttataatgct 60
 ccaccctaca gaaatgggac cccaagtacc cgatcttccc tttaggagag gcaggcaggt 120
 gggtagcag cagatgtagt ttccatttcc ctgggggttt aattttccaa actttgtctt 180
 tt 182

<210> 93
 <211> 1783
 <212> DNA
 <213> Homo sapien

<400> 93
 ccgccctgca ccctccaagc ctgcacctct tagtccagca ccaggaagag tggctctagg 60
 agggagagga gccaggggac tgacccaaga tgccgttgac tccagagccg ccctctgggc 120
 gcgtggaggg gcccccgca tgggaagcag ccccatggcc ctcactgcc tgtgggccct 180
 gcatcccat catgctggtc ctggccaccc tggctgcgct cttcatcctc accaccgctg 240
 tgttggtga acgcctgttc cgccgtgctc tccgccaga cccagccac cgtgcaccca 300
 ccctggtgtg gcgcccagga ggagagctgt ggattgagcc catgggcacc gccgagagc 360
 gctctgagga ctggtatggc tctgcggtcc ccctgctgac agatcgggcc cctgagcctc 420
 ccaccaggt gggcactttg gagggccgag caacagcccc acctgcccc tcagcccca 480

65

```

attctgctcc cagcaacttg ggccccaga ccgtactgga ggtcccagcc cggagcacct 540
tctggggggc ccagccctgg gaggggaggg ccccgccaca ggcttgggtga gctgggctga 600
acccgagcag aggccagagg ccagcgtcca gtttgggagc cccagtgcc aggaggcagt 660
ggccaggagc ccggatcctg agtggggcct ccagccacgg gtcaccttgg agcagatctc 720
agctttctgg aagcgtgaag gccggaccag tgtggggttc tgaatcccca gggttcccca 780
gagacccccg aggcaggcct tgcctcagtg ggaccgggga cccagggatc cagcattagg 840
attgagtctg cccagcgtg gatgcccttc ccaggctcct tccacctggg agtccccctc 900
cccggtctg tgttgttgtc ctggtctatg tggatctagg ggtaagcca tgcagtgcct 960
ctgctcagct acctgggctg tggctcagag tacctggggg tggagccaat gccaggccag 1020
aagccttcaa gatcgcatcc agatgaagaa ccctaggtat ctagatagtc aggaaatggc 1080
atcgaccagc cacctccacc ttctttcagt gtttaccgaa gccaccaata ccaaagagaa 1140
cgggtcctgc ggtgctgaac agcctcgggtg tggcgatgac agctggcagg agatgacagg 1200
aatccagttt cccagagcca caaatcctgt tctccttggc cactcaccca ctgtgaggtc 1260
ctctaggaaa atacacaaag agaggaccag accaggcaga ggaacatttt gtttcatatg 1320
aactgtggct ttgaccccca aactgcaagg aggaacttgc tgggccaagc tgcagcggcg 1380
ctgtcttgct ggagtgggga cctagagtca gagaaaaccc acaggctcct ctgcccattc 1440
tcctccatct gcacacgtct cagcctcggg ccctcaccac tccatgggtga ggaaggccat 1500
ggccagggga aactgagttt catccaatgt ggagaggagc gttgtcctag agcagggcaa 1560
ctcccaaact gtgacctctg atcatcgtcc cttccagctt gctggagtgt ccagagagac 1620
agatttgcca caagctaggc ttacttataa tgctccacc tacagaaatg ggaccccaag 1680
tacccaatct tccctttagg agaggcaggc aggtgggtga gcagcagatg tagtttccat 1740
ttccctgggg gtttaatttt ccaaacttgt cttttttttt ttt 1783

```

<210> 94

<211> 612

<212> DNA

<213> Homo sapien

<400> 94

```

gtgtaaaata agtctcatga tatttgcctc tcatggtagt tataaagatt gagattctgt 60
gtgaaaggca cctggcacac aggtgctcac taaatgcagt tcctttcctc tgctgtactt 120
agacattttc taggaaggat aggtccaata ggactgaata ttttathtag aagagatggc 180
ttcctttaag ttgccaaca agaggtttgg aagaaaacca aatgcttacc agccacttca 240
tattaatacc agtcattttc tcccttcagt atcaatgcct aggagccagg aagctgtgcc 300

```

66

```

aatgccagtg gttgtggaga tggcaaaaga agggaggcca gccacctggg acagctgaat 360
ccaagcctga ctctcagccc cagaagggtg gacaagatgc tgccaattca tcgaacccaa 420
agaaagctgc agagatcact gttatccagc agacatactt ctgaaaagtt ctgctctatc 480
tcaaagactg aatgatacta cacagtcctc tccctattaa tatgggcaca ttcttgccag 540
tttcacactt gtatcttcag caggacatc aatcaaacac caattcctgg ttaatgccgg 600
gagagtgtgg gc 612

```

```

<210> 95
<211> 1839
<212> DNA
<213> Homo sapien

```

```

<400> 95
cttgggtcca gctgcgtgtg gtgaaagcaa ctagaggcag agctatcaag ggctgtgaca 60
gatgagcagt ggtctgtctg caatgagcat gtgctcaagc taacatggat accatcttgg 120
tcttcagcct aatcattgca tcctatgatg ccaacaagaa agacctcaga gatagcagtt 180
gccgactgga acagctgcct gggatcttcc caaaagacgt gagaagcatc agagaattgc 240
aaatgcaaga aactcacaca gaaacaaaaa ggacaacatt cattcaaaac cggactatag 300
ctacctgca gtgccttggc tctgacagca aagtaaaagt caaccttgta tatttggaga 360
gaaggccaaa ggtcaagcat attttgaaga acctgagaat cattgctgct ccccgagaa 420
acagctctgc ctctcaagc tgtcacctaa tccccacatc caagtttcag actggatctc 480
ttctaaaagg caaagtatca atgcctagga gccaggaagc tgtgccaatg ccagtgggtg 540
tgagatggc aaaagaaggg aggccagcca cctgggacag ctgaatcaa gcctgactct 600
cagccccaga aggttgga caatgctgcc aattcatcga acccaaagaa agctgcagag 660
atcactgtta tccagcagac atacttctga aaagtctgc tctatctcaa agactgaatg 720
atactacaca gtctctccc tattaatatg ggcacattct tgccaatttc acaattgtat 780
cttcagcagg gacattacaa tcaaacacca attcctggtt aatgaaggga gagtgtgggc 840
ttagcagagt taccctcatg cccctatctg agccacaacc ttctgtaatt cacttcatac 900
atccatctaa atggatacct ttccatcccc tcaaacgaga acaaaaagta ttccctgcaa 960
gcactatgaa tggaccttac tactctcttt gacagaagac tcaaacacag cctctaagaa 1020
acaaggcagc tgtagtgtg acatagcttc cagctcctct gttatcttca cagctgactt 1080
cacactcact ggccctcaat agcttagagt gggactcctg atctccctgg ttgacgttta 1140
taagggtttg taatttagct tctggtacta catgtatcta tccacacact gctgaaggga 1200
acttgacct acgcattcca aaagccttcc tcataagtta tccccagggt gttcagggac 1260

```

67

```

tacagaccat tagtacccaa gtgaagtctt ctgacgatcc agaattccta aggtggctct 1320
gatattatat catactttta tagagagttg catacctttt ggacaaagat tccagagcaa 1380
aataattatg ataatgctgt ttctcacagg agaatagcta tgagtcaaat ctaacctgct 1440
taaataagag tttggtgaga aagtgaacc acctgatctt gaatcaatgt tgaggtgaaa 1500
aaggaaatgt caggagggat aagacagggt gaggccttgc ttcttttctt aagagtctga 1560
aaccattcca ttctattttg gtgaaatgtg tctgtttcta agaaattctg ttttcttagt 1620
tctctgggta aaaaaaaaaa tacataaact gtatccttct ttcttttact tcaagtgtctg 1680
tccaaaaagt gattatcaaa acaataccaa caggggaaaa tctcaccta aggctcgatt 1740
taatattcaa gtccagcctg aaagagaaca catataaagc atggataatg aggatatgtt 1800
tcaaagaagc tccaaagaac ctctagggcc aggcattggt 1839

```

```

<210> 96
<211> 5043
<212> DNA
<213> Homo sapien

```

```

<400> 96
atggcagaag gcaaggagaa gcaagtcaca tcttacatgg atggcagcag accttatgac 60
gtgagcatga cctacatcca caaagcagga gggcctgatc agcaggaatt agtcatgtct 120
acctgcactg tgccccctga ctctgttgt catctgcctc aggcccgaa gaattatcgc 180
aaatatttcc ggtcagaggc tgccttcacc ttggcagatt tcatctacaa gagcatgatc 240
cgagtaaata gttccagatt ggttcgagtc acccaggtgg agaagagga gaaactgaag 300
gagctagagc agtttagtat ctggaacttt ttttctcctt ttttaaaaga gaaattgaat 360
gacacctatg ttaacgtggg tctatacagc acaaaaacct gcctcaaagt tgagattata 420
gagaaggaca ccaagtacag tgtcattgtg atccggaggt cctgggatgt tatccgagta 480
aatagttcca gattggttcg agtcacccag gtggagaatg aggagaaact gaaggagcta 540
gagcagttta gtatctggaa ctttttttcc tcctttttta aagagaaatt gaatgacacc 600
tatgttaacg tgggtctata cagcacaaaa acctgcctca aagttgagat tatagagaag 660
gacaccaagt acagtgtcat tgtgatccgg aggtcctggg atgttatccg agtaaatagt 720
tccagattgg ttcgagtcac ccagggtggag aatgaggaga aactgaagga gctagagcag 780
tttagtatct ggaacttttt ttctcctttt ttaaaagaga aattgaatga cacctatgtt 840
aacggcatac cgtggacaaa ggtggactac tttgataatg gcatcatttg taagctcatt 900
gagcataatc agcgaggat cctggccatg ttggatgagg agtgcctgcg gcctggggtg 960
gtcagtgact ccacttttct agcaaagctg aaccagctct tctccaagca tggccactac 1020

```

| | |
|---|------|
| gagagcaaag tcacccagaa tgcccagcgt cagtatgacc acaccatggg cctcagctgc | 1080 |
| ttccgcatct gccactatgc gggcaagggtg acatacaacg tgaccagctt tattgacaag | 1140 |
| aataatgacc tactcttccg agacctgttg caggccatgt ggaaggccca gcacccccctc | 1200 |
| cttcgggtcct tgtttcctga gggcaatcct aagcaggcat ctctcaaacg cccccgact | 1260 |
| gctggggccc agttcaagag ttctgtggcc atcctcatga agaactctgt ttccaagagc | 1320 |
| cccaactaca tcagggtgcat aaagcccaat gagcatcagc agcgagggtca gttctcttca | 1380 |
| gacctgggtg caacccaggc tcggtacctg ggactgctgg agaacgtacg ggtgcgacgg | 1440 |
| gcaggctatg cccaccgcca gggttatggg cccttcctgg aaaggtagcg attgctgagc | 1500 |
| cggagcacct ggctcactg gaatggggga gaccgggaag gtgttgagaa ggtcctgggg | 1560 |
| gagctgagca tgtcctcggg ggagctggcc tttggcaaga caaagatctt cattagaagc | 1620 |
| cccaagactc ttttctacct cgaagaacag aggcgcctga gactccagca gctggccaca | 1680 |
| ctcatacaga agattttaccg aggtggcgc tgccgcaccc actaccaact gatgcgaaag | 1740 |
| agtcagatcc tcatctcctc ttggtttcgg ggaacatgg cccgaaagaa ttatcgcaaa | 1800 |
| tatttccggg cagaggctgc cctcaccttg gcagatttca tctacaagag catggtacag | 1860 |
| aaattcctac tggggctgaa gaacaatttg ccatccacaa acgtcttaga caagacatgg | 1920 |
| ccagccgccc cctacaagtg cctcagcaca gcaaactcagg agctgcagca gctcttctac | 1980 |
| cagtgggaagg caactcctgt ccctccttcc tcacagtgca agaggttccg ggatcagctg | 2040 |
| tccccgaagc aggtagagat cctgagggaa aagctctgtg ccagtgaact gttcaagggc | 2100 |
| aagaaggctt catatcccca gagtgtcccc attccattct gtggtgacta cattgggctg | 2160 |
| caaggggaacc ccaagctgca gaagctgaaa ggcggggagg aggggcctgt tctgatggca | 2220 |
| gaggccgtga agaagggtcaa tcgtggcaat ggcaagactt cttctcggat tctcctcctg | 2280 |
| accaagggcc atgtgattct cacagacacc aagaagtccc aggccaaaat tgtcattggg | 2340 |
| ctagacaatg tggctggggg gtcagtcacc agcctcaagg atgggctctt tagcttgcat | 2400 |
| ctgagtgaga tgtcatcggg gggctccaag ggggacttcc tgctggtcag cgagcatgtg | 2460 |
| attgaactgc tgacaaaaat gtaccgggct gtgctggatg ccacgcagag gcagcttaca | 2520 |
| gtcaccgtga ctgagaagtt ctcaagtggg ttcaaggaga acagtgtggc tgtcaaggtc | 2580 |
| gtccagggcc ctgcagggtg tgacaacagc aagctacgct acaaaaaaaaaa ggggagtcac | 2640 |
| tgcttgaggg tgactgtgca gcagctcacg gccggctatc atgctggcca gggagaactg | 2700 |
| attaacttct ctctctgcct gcagattaat ctgctgtctg agcacaagcc acgtgcttct | 2760 |
| ggcacaccct gctttgagct gagatagaac ctggggaatc atctgttttc aggcgggtga | 2820 |

| | |
|---|------|
| ggggctagag cctgccttgt ttgggaggag ggtggctctg ttcagaatag gggtagctca | 2880 |
| ggctctggcc agccttctcc cgcccccaac agctcccccc atccttgact tctcagaatc | 2940 |
| aggccgagaa gagcctatct ggccgagagt ggggtggtga cctgcgcctc atcgcccccg | 3000 |
| ctctccatct catctcctgc tcccagggcc caaattgtcg tcactttccc agtgaagtgt | 3060 |
| ctggtcattt tcagaagcaa tttcaggaga acatgcagct gccgctccct atcctgcatt | 3120 |
| tcccttcaca gggctgaagg cactgtcagc tccctgggct gggggtgatg ggagagggga | 3180 |
| agggtaggg cccaccccc tgtctcact gtgccatca tgtagatgga ctggagttca | 3240 |
| aggaagggca ggcactcccc tctccttta ctctctgtc actctcttc tctctctctt | 3300 |
| tctgtctct gcctctctt tctggagcct aggagtgtgt gttttcatcc cctgaaacaa | 3360 |
| atagggactc agtttcccca cctgtgttac agggttggaa ttggctccat cactgtggga | 3420 |
| gaagctggag ttctgtacc agtcctcccc tccccagccc tgctctctct cctccagccc | 3480 |
| tctcccttca gccagttcag cgctctgaga gtctgggttg tttcagcctc tgaggggcac | 3540 |
| aagccatcct ggattccct aaccccatga ggagccattc tagcatctca cagcttaaac | 3600 |
| cagctctagc tcagtcctcc tggcttagtc catTTTTctt cctcaggctc tgagggcctc | 3660 |
| ttgttccttg ctctgtgggg tttctccag ttgtctcctg gctgcaggac atggcaggac | 3720 |
| atagaatgct gtcaccttc cactcttcat tggcatctcc acccagtgtc acatatgacc | 3780 |
| ctagccgtgc tctcccttg ccagtacccc tctgggattt tgcgagagtc cacaagttgt | 3840 |
| gcatgtggtg gatatttca ggccatcttg tgtgtacaag ctagagggtc tgcttccacc | 3900 |
| tctggccctc agtgaattgc tgactaacct gtctcaacac agcacaactg tacacacctt | 3960 |
| ttcctggcct catccctaac ccatcatagc agcaaagagg ggaagttgca ggggaggagc | 4020 |
| tgctaaggac cctggactcc aagtacctg ctctctagc ccaggacat catctgagat | 4080 |
| gtggctcaaa taaaggggtg gtgttcaaga aaaaacactt ggggactcta tagctgcaac | 4140 |
| accacttta catgtcattt ccatatgatt tgtaggcaaa atgaagcca ggctgtccta | 4200 |
| gccctccaat acctcctct ctcatcacct ctccaacata gcctagcatt agctctttca | 4260 |
| agtctttgct aatcccagag atcaaggggt gatcaactct cctgccatc ccttgttcc | 4320 |
| ccgcaccccc cgccccggct cccccacat ccttggctcc tgccatcctc tttgagatgc | 4380 |
| tgcatcatca aaggacatta tttatggtgt acctttgctg aagccctgct tccctggtgc | 4440 |
| cagggcttgg gagcagggat ggggtgggtg gtgggggaga ggggtggatg cagagattgg | 4500 |
| accaggagg cttttagtcc tcagctcttg gcttaacacc tctcctctt acacacccaa | 4560 |
| ctcctccag cctgcccagc ttgggccttc agctccagat tgggtggggtt aggagaggag | 4620 |
| gaggagggag atggatggaa ccaattagga acagcacctg ggctcctcac aggaatgaac | 4680 |

cagtcatgcc atttgcatgt aaacagcttc ccacttctct cctcatccta ccaaagtctc 4740
 ccaaccctgg gttctggccc atgttctttg cccacacagc cctgtaatta gctgggtaat 4800
 gagaagcttt taatgagtc ctagcatc tcgtgtaata aagaggcctt gagaccagc 4860
 tgctgtctc actttgggat gaacacgggt ccctgtgtag ccagtgactt ctgtcagtac 4920
 agtctaagtt ctggatggg gtgggagaca aacatttcag gacccagca gcacttgaga 4980
 ggttccatgg tggatccatg tttttgactg tgatacaaga aacttggtc tggcttcctt 5040
 gtt 5043

<210> 97
 <211> 1171
 <212> DNA
 <213> Homo sapien

<400> 97
 gggattgtca tctttagatc ctgactttcc aattccatct tggatgtgct acccgtggcc 60
 agaagccact ggggtgtgcc ttactttgac ctttcaccag tacattccat cttagggggc 120
 gtttctctaa gttggcttca gactccctcc tgacccttta taccaccca aaaccaaagg 180
 ccctggttta ggagctgaat tgtagcagcc gtccctcagt atctagatcg cagggagatt 240
 tcaactggagc tggctgaagc tgttggcgct ttgagcagcg ttgagccatc agtcctagtt 300
 ctgctctttt gcccactgc agcttttggc caettgcccc agtttagtgt caccaggccc 360
 cgactataag ccacagtaat ttttatggct gcagggtatta gagggcagcc agtctacaga 420
 ggtaggggct gcccatcctc cagctgtctt ttgcaatacg ggcatacact tcgagccaac 480
 agcagctcca gccggccgtg ccagctgtca aatgctttct gggctggcac ggctctttga 540
 atccgtggta atggcttctt ctgttgaatg tcgtccaaaa gcaattgctt gtcagcttaa 600
 aagcgtgtgc catgttgtca tggccgctg ctgagagggc aggcacccc gcacacttgc 660
 agactgggat gctgtagcat cttctcatga gtcagggatt ggggtgtaact tttgcaaaa 720
 tttctgcat tttccccacc ctctccaaca tgtctttgaa taccatata cctttacatt 780
 tgagatgaaa gtggaactct aaaaaccttg gaaagattct cttgtgaggt gacttctact 840
 taagcttggc cacagcctcc tgctgtctc aaagagtggg caggagattt cttcccttgg 900
 acgctgccag tgctggaata tgaaacttac caagtgtcag cagttggggc ctgtccttca 960
 cggtgaccca gcccctcat tgggtgtctg atgattgtgg acaccagtc agccccattt 1020
 tctggccaca ccaagcatgg caatgacacc aggttggcac aaaacaactc tgccctggatt 1080
 gattacttag cccatcatag cgtacacttt cttcatttta gtaagttgtg gttttatttt 1140
 gaataagtct gtttcaacct attaatatta g 1171

<210> 98
 <211> 660
 <212> DNA
 <213> Homo sapien

<400> 98
 taacatgctg aatggggttg ctctcataga aggtattctt gaggggcag ttttccttac 60
 aaaacatagt atactaagac tgtggtattc agacagctct tcaatcttca agggaaactgt 120
 ggaagtaaac caggggtgac atggattatc aaatattgtg gatcagggtga gctctaggct 180
 tetcaatgca aggtaattac cctgcacttt catctatttt ataattattgc ctttattttt 240
 ttccatgaaa gagtaagggt tggaaaaaca agaaaaacac gggttaaaatt atttctctag 300
 tgaattgggt gtgtttataa atcttttagct tagcttgagt ccccttttgt gaaatgtata 360
 aatccttttg atacattaac aaacacttggt aaaatgactt aacttattaa gttgaattct 420
 ctaatatag gggtataagt gaaggcttaa ttattcagca ttactgaatt caatatatct 480
 gccctttcag tttgcctagg gtgacaaata catttcccat tacagcttgc caatgtaaca 540
 cagccttatt tcatttttgt tgtaaataga atcacattct aatatgtttt agctgaaaac 600
 agcaaagtat atattcaact acttcccaat agggaaaatt tttaaaaaaa gttatttctt 660

<210> 99
 <211> 3002
 <212> DNA
 <213> Homo sapien

<400> 99
 ggaatctggc ccctagaggc tggacttggt gcccgaaacc cccatctccg gcggagagac 60
 cgtccgaggt aattgtctgc cagcagtgca cattctgaaa acaggactgt ggaaactgaa 120
 gtcattgcag agtttggatc tgtcattcaa tgggatattg caaatagggt ggtctgattt 180
 tcacaactgc ctgcaactgg agaattctctg tttaaagagc aacaagatat tcaaaattcc 240
 cccacaagcc ttcaaggacc tcaaaaaatt acaggtcata gaccttagca acaatgctct 300
 gattaccatc ctaccaatga tgatcatagc tctagaattt ccccatctag tgggtgactt 360
 ggctgataat aactggcagt gtgatgatag tgtggcagtc tttcaaaatt ttatttctga 420
 atcctggagg aaaaagtgga atgtcatttg caacaggctc atagggagtg aggaggccaa 480
 cgggggcact cccagagca ggatttccag ggaaaccgc cttcctccca ttcattctgca 540
 tcgcatgaaa agcctcataa ggagcaaagc agagaggccc caggaggaa ggcacacggg 600
 catttctact ctggggaaga aggcaaaggc cggctctggt ctcaggaaga agcagagacg 660
 gctgccaagg agtggttagaa gcaccgcga tgtgcaggct gccggcaaaa aagaggacgc 720

| | | | | | | |
|------------|------------|------------|-------------|------------|------------|------|
| tccccaggac | ctggctcttg | cggtgtgcct | gtcagtggtc | atcacattcc | ttgtcgcctt | 780 |
| cagcctgggg | gctttcacaa | ggccttatgt | tgacagactg | tggcaaaaaa | agtgccagag | 840 |
| caaaagccct | ggcctggaca | acgcgtattc | aaacgagggc | ttctacgatg | acatggaagc | 900 |
| tgcggggcac | acaccacacc | cagagaccca | tctgcgccaa | gtatttcctc | atctaagcct | 960 |
| ctacgagaac | cagacccctt | tctgggtgac | acagccacac | ccacacgcc | ccgtaattcc | 1020 |
| tgatagaact | ctgggaagga | gcagaaagga | tcctggcagt | tcgcagagcc | caggacagtg | 1080 |
| cggggacaac | accggggcag | gaagtggaaa | tgatgggtgca | gtctattcca | ttctccagag | 1140 |
| acatccacat | gccggtaac | gtgaactaat | gtcagcagcg | caggaccaca | tccataggaa | 1200 |
| tgatattctc | ggagaatgga | cttatgaaac | tgtggcccag | gaagagcctc | tcagtgcaca | 1260 |
| ttcagtgggc | gtctcttctg | tagctggcac | gtctcacgct | gtctctggct | caagccgtta | 1320 |
| tgattccaat | gaattagacc | ttcccctctc | cggagaaata | acagcttccc | tctgtaaaat | 1380 |
| gctaacacat | gcagaagcac | agaggactgg | agatagtaag | gaaagagggg | gcactgaaca | 1440 |
| gtcacttttg | gactcgcaga | tggaaatttc | taaggaaagg | caagtgagtt | catccattga | 1500 |
| tttgctgagc | atacagcagc | caaggctgtc | cggggcaagg | gctgaggaag | cgctttcagc | 1560 |
| ccactacagc | gaggttccat | acggtgaccc | aagagacaca | ggcccatcag | tctttcctcc | 1620 |
| aagatgggac | agtggccttg | atgtcactcc | tgctaacaag | gaaccagtgc | agaaatccac | 1680 |
| tcttctgac | acttgctgtg | agttggagag | tgactgtgac | tctgatgagg | ggctctgttt | 1740 |
| cactctgagc | tccataagtt | cagagagtgc | aaggagcaag | actgaagagg | cagtgcctga | 1800 |
| tgaggagtcc | ctgcaggacg | agagctcagg | ggcaagcaag | gacaatgtga | cggctgtaga | 1860 |
| cagtcttgag | gaaaatgtta | ccttccaaac | aattccaggg | aaatgcaaga | atcaagaaga | 1920 |
| tccctttgaa | aaacctctca | tttctgctcc | agactctggc | atgtacaaga | ctcatctgga | 1980 |
| aaatgcctct | gacactgata | gatctgaggg | cctgtcaccc | tggcccaggt | caccagggaa | 2040 |
| tagtccctta | ggggatgagt | ttccgggcat | gttcacttat | gattatgaca | cagctcttca | 2100 |
| atccaaggca | gcagaatggc | attgctcact | tagagactta | gaattttcaa | atgtggacgt | 2160 |
| tttacagcaa | acaccacat | gttctgctga | agttccctca | gatcctgata | aggctgcctt | 2220 |
| ccatgaaaga | gactcagaca | ttttaaata | agaatctttc | attaaggaaa | tattcacagc | 2280 |
| tctaaacaat | attcttttaa | agatcatggc | aggggaaaac | taaagccttc | acaacaagat | 2340 |
| cctgaaggag | acaacatgaa | ttccaaccaa | gtggacactg | atgcaaatga | gggttttgtg | 2400 |
| ggcccacttg | agggcgatga | ttctagaagg | gttataagct | aaacacactt | gttaacatcc | 2460 |
| tatggggatg | aacctgctct | tcagtgtgaa | agagaaagag | gggaatattt | tgagatcatt | 2520 |
| ccaagaacca | tgcaccatta | ttacaagagc | tccaaacaaa | accagttcac | tgagaacaca | 2580 |

```

ggagcacttc atcaataggt tggggtaaat attcagaaga gaatctgttg cagtttagacg 2640
aagtggttct aactttcctt ctcaaataca gacctggagc aatctgatgg gatttggtctc 2700
tctgaaaaga acagcttttag gatgacaaag acaaggatgt tcaacttgat catcaaagag 2760
aattcaaatg acattttttc atacatattg aagcacttaa gtggaaatgc tggatagatt 2820
ttaatgaatt ctgaacacac actagcacia ctttttggag gctgatcgtg aaactgcaaa 2880
attgaaaaag ttctgagtaa atataaatgt ttttaaataa gttatgttaa aagagaagag 2940
gggccacttt ttagaataaa gtgtaaatta aaaatcaaaa aaaaaaaaaa aaaactgcgg 3000
cc 3002

```

```

<210> 100
<211> 312
<212> DNA
<213> Homo sapien

```

```

<400> 100
attcacagct ctaaacaata tttccttaat gaaagattct tattttaaaa tgtctgagtc 60
tcagatcctg ataaggctgc cttccatgaa agagactcag acattttaaa ataagaatct 120
ttcatggaag gcagccttat caggatctga gggaacttca gcagaacatg gtggtgtttg 180
ctgtaaaaag tccaccagga tctgagggaa cttcagcaga acatgggtgt gtttgctgta 240
aaacgtccac atttgaaaat tctaagtctc taagtgagca atgccattct gctgccttgg 300
attgaagagc tg 312

```

```

<210> 101
<211> 59
<212> PRT
<213> Homo sapien

```

```

<220>
<221> MISC_FEATURE
<222> (6)..(6)
<223> x=any amino acid

```

```

<400> 101

```

```

Met His Ala Thr Pro Xaa Leu Thr Leu Pro Phe Leu Ser Tyr Ile Tyr
1           5           10           15

```

```

Met Val Leu Pro Phe Leu Pro Pro Tyr Ser Ser Ser Arg Asp Ser Ser
20           25           30

```

```

Tyr Leu Glu Cys Ser Ser Cys His Ile Cys Met Ile Lys Ser Cys Thr
35           40           45

```

74

Arg Leu Gln Thr His Leu Gly Tyr His Leu Val
 50 55

<210> 102
 <211> 64
 <212> PRT
 <213> Homo sapien

<400> 102

Met Ser Thr Tyr Asn Ile Phe Leu Ile Ile Met Pro Ile Ser Phe Ser
 1 5 10 15

Lys Lys Asn Val His Asn Met Glu Tyr Ile Phe Pro Asp Leu Phe Asn
 20 25 30

Ser Leu Met Asn Lys Thr Asp Ile Phe Tyr Lys Cys Asp His Ser Val
 35 40 45

His Thr Trp Leu Lys Gln Leu Tyr Met Thr Phe Ile Ile Arg Lys Asn
 50 55 60

<210> 103
 <211> 49
 <212> PRT
 <213> Homo sapien

<400> 103

Met Val Thr Ala Thr Ile Tyr Pro Ile Met Asp Ala Ile Cys Phe Arg
 1 5 10 15

Ala Arg Glu Asn Ala Lys Asp Lys Ser Arg Leu Asn Glu Gln Leu Ala
 20 25 30

Gln Ile Pro Ile Ser Leu Met Pro Leu Ser Gln Leu Ile His Thr Val
 35 40 45

Ser

<210> 104
 <211> 40
 <212> PRT
 <213> Homo sapien

<400> 104

Met Lys Ile Pro Tyr Leu Lys Gly Lys Tyr Ser Leu Ile Leu Asn Cys
 1 5 10 15

75

Asn Ala Gly Lys Pro Asn Cys Phe Gly Ile Cys Asn Leu Asn Val Leu
 20 25 30

Gln Asn Leu Val Leu Lys Phe Asp
 35 40

<210> 105
 <211> 51
 <212> PRT
 <213> Homo sapien

<400> 105

Met Ile Asn Asp Pro Lys Leu Leu Tyr Leu Ser Asn Pro Cys Val Pro
 1 5 10 15

Phe Leu Leu Phe Lys Lys Thr Ile Ser Pro Cys Arg Cys Leu Ser Leu
 20 25 30

Phe Cys Phe Cys Val Ser Val Leu Ser Tyr Ile Phe Ser Pro Pro Ile
 35 40 45

Cys Cys Phe
 50

<210> 106
 <211> 45
 <212> PRT
 <213> Homo sapien

<400> 106

Met Gln Arg Tyr Thr Leu Ser Thr Ile Gln Tyr Val Phe Pro Ala Phe
 1 5 10 15

Leu Gly Ile Tyr Ile Ala Val Ala Ile Phe Leu Ala Lys Ile Gly Ser
 20 25 30

Tyr Tyr Ser Phe Pro Ile Leu Phe Phe Lys Val Asn Phe
 35 40 45

<210> 107
 <211> 51
 <212> PRT
 <213> Homo sapien

<400> 107

Met His Asp Phe Leu Arg Met Ser Leu Pro Ser Asp Val Gly Ser Asp
 1 5 10 15

76

Glu Thr Tyr Ser His Lys Val Met Ser Ser Thr Met Leu Ser Leu Lys
 20 25 30

Ala Phe Glu Glu Leu Gly Gly Gln Val Cys Trp Gln Arg Pro Val Ile
 35 40 45

Pro Ala Val
 50

<210> 108
 <211> 928
 <212> PRT
 <213> Homo sapien

<400> 108

Met Ala Glu Gly Lys Glu Lys Gln Val Thr Ser Tyr Met Asp Gly Ser
 1 5 10 15

Arg Pro Tyr Asp Val Ser Met Thr Tyr Ile His Lys Ala Gly Gly Pro
 20 25 30

Asp Gln Gln Glu Leu Val Met Leu Thr Cys Thr Val Pro Leu Asp Ser
 35 40 45

Cys Cys His Leu Pro Gln Ala Arg Thr Asn Tyr Arg Lys Tyr Phe Arg
 50 55 60

Ser Glu Ala Ala Phe Thr Leu Ala Asp Phe Ile Tyr Lys Ser Met Ile
 65 70 75 80

Arg Val Asn Ser Ser Arg Leu Val Arg Val Thr Gln Val Glu Asn Glu
 85 90 95

Glu Lys Leu Lys Glu Leu Glu Gln Phe Ser Ile Trp Asn Phe Phe Ser
 100 105 110

Ser Phe Leu Lys Glu Lys Leu Asn Asp Thr Tyr Val Asn Val Gly Leu
 115 120 125

Tyr Ser Thr Lys Thr Cys Leu Lys Val Glu Ile Ile Glu Lys Asp Thr
 130 135 140

Lys Tyr Ser Val Ile Val Ile Arg Arg Ser Trp Asp Val Ile Arg Val
 145 150 155 160

Asn Ser Ser Arg Leu Val Arg Val Thr Gln Val Glu Asn Glu Glu Lys

77

| | | |
|---|-----|-----|
| 165 | 170 | 175 |
| Leu Lys Glu Leu Glu Gln Phe Ser Ile Trp Asn Phe Phe Ser Ser Phe | | |
| 180 | 185 | 190 |
| Leu Lys Glu Lys Leu Asn Asp Thr Tyr Val Asn Val Gly Leu Tyr Ser | | |
| 195 | 200 | 205 |
| Thr Lys Thr Cys Leu Lys Val Glu Ile Ile Glu Lys Asp Thr Lys Tyr | | |
| 210 | 215 | 220 |
| Ser Val Ile Val Ile Arg Arg Ser Trp Asp Val Ile Arg Val Asn Ser | | |
| 225 | 230 | 235 |
| Ser Arg Leu Val Arg Val Thr Gln Val Glu Asn Glu Glu Lys Leu Lys | | |
| 245 | 250 | 255 |
| Glu Leu Glu Gln Phe Ser Ile Trp Asn Phe Phe Ser Ser Phe Leu Lys | | |
| 260 | 265 | 270 |
| Glu Lys Leu Asn Asp Thr Tyr Val Asn Gly Ile Pro Trp Thr Lys Val | | |
| 275 | 280 | 285 |
| Asp Tyr Phe Asp Asn Gly Ile Ile Cys Lys Leu Ile Glu His Asn Gln | | |
| 290 | 295 | 300 |
| Arg Gly Ile Leu Ala Met Leu Asp Glu Glu Cys Leu Arg Pro Gly Val | | |
| 305 | 310 | 315 |
| Val Ser Asp Ser Thr Phe Leu Ala Lys Leu Asn Gln Leu Phe Ser Lys | | |
| 325 | 330 | 335 |
| His Gly His Tyr Glu Ser Lys Val Thr Gln Asn Ala Gln Arg Gln Tyr | | |
| 340 | 345 | 350 |
| Asp His Thr Met Gly Leu Ser Cys Phe Arg Ile Cys His Tyr Ala Gly | | |
| 355 | 360 | 365 |
| Lys Val Thr Tyr Asn Val Thr Ser Phe Ile Asp Lys Asn Asn Asp Leu | | |
| 370 | 375 | 380 |
| Leu Phe Arg Asp Leu Leu Gln Ala Met Trp Lys Ala Gln His Pro Leu | | |
| 385 | 390 | 395 |
| Leu Arg Ser Leu Phe Pro Glu Gly Asn Pro Lys Gln Ala Ser Leu Lys | | |
| 405 | 410 | 415 |

Arg Pro Pro Thr Ala Gly Ala Gln Phe Lys Ser Ser Val Ala Ile Leu
 420 425 430

Met Lys Asn Leu Tyr Ser Lys Ser Pro Asn Tyr Ile Arg Cys Ile Lys
 435 440 445

Pro Asn Glu His Gln Gln Arg Gly Gln Phe Ser Ser Asp Leu Val Ala
 450 455 460

Thr Gln Ala Arg Tyr Leu Gly Leu Leu Glu Asn Val Arg Val Arg Arg
 465 470 475 480

Ala Gly Tyr Ala His Arg Gln Gly Tyr Gly Pro Phe Leu Glu Arg Tyr
 485 490 495

Arg Leu Leu Ser Arg Ser Thr Trp Pro His Trp Asn Gly Gly Asp Arg
 500 505 510

Glu Gly Val Glu Lys Val Leu Gly Glu Leu Ser Met Ser Ser Gly Glu
 515 520 525

Leu Ala Phe Gly Lys Thr Lys Ile Phe Ile Arg Ser Pro Lys Thr Leu
 530 535 540

Phe Tyr Leu Glu Glu Gln Arg Arg Leu Arg Leu Gln Gln Leu Ala Thr
 545 550 555 560

Leu Ile Gln Lys Ile Tyr Arg Gly Trp Arg Cys Arg Thr His Tyr Gln
 565 570 575

Leu Met Arg Lys Ser Gln Ile Leu Ile Ser Ser Trp Phe Arg Gly Asn
 580 585 590

Met Ala Arg Lys Asn Tyr Arg Lys Tyr Phe Arg Ser Glu Ala Ala Leu
 595 600 605

Thr Leu Ala Asp Phe Ile Tyr Lys Ser Met Val Gln Lys Phe Leu Leu
 610 615 620

Gly Leu Lys Asn Asn Leu Pro Ser Thr Asn Val Leu Asp Lys Thr Trp
 625 630 635 640

Pro Ala Ala Pro Tyr Lys Cys Leu Ser Thr Ala Asn Gln Glu Leu Gln
 645 650 655

Gln Leu Phe Tyr Gln Trp Lys Ala Thr Pro Val Pro Pro Ser Ser Gln
 660 665 670

Cys Lys Arg Phe Arg Asp Gln Leu Ser Pro Lys Gln Val Glu Ile Leu
 675 680 685

Arg Glu Lys Leu Cys Ala Ser Glu Leu Phe Lys Gly Lys Lys Ala Ser
 690 695 700

Tyr Pro Gln Ser Val Pro Ile Pro Phe Cys Gly Asp Tyr Ile Gly Leu
 705 710 715 720

Gln Gly Asn Pro Lys Leu Gln Lys Leu Lys Gly Gly Glu Glu Gly Pro
 725 730 735

Val Leu Met Ala Glu Ala Val Lys Lys Val Asn Arg Gly Asn Gly Lys
 740 745 750

Thr Ser Ser Arg Ile Leu Leu Leu Thr Lys Gly His Val Ile Leu Thr
 755 760 765

Asp Thr Lys Lys Ser Gln Ala Lys Ile Val Ile Gly Leu Asp Asn Val
 770 775 780

Ala Gly Val Ser Val Thr Ser Leu Lys Asp Gly Leu Phe Ser Leu His
 785 790 795 800

Leu Ser Glu Met Ser Ser Val Gly Ser Lys Gly Asp Phe Leu Leu Val
 805 810 815

Ser Glu His Val Ile Glu Leu Leu Thr Lys Met Tyr Arg Ala Val Leu
 820 825 830

Asp Ala Thr Gln Arg Gln Leu Thr Val Thr Val Thr Glu Lys Phe Ser
 835 840 845

Val Arg Phe Lys Glu Asn Ser Val Ala Val Lys Val Val Gln Gly Pro
 850 855 860

Ala Gly Gly Asp Asn Ser Lys Leu Arg Tyr Lys Lys Lys Gly Ser His
 865 870 875 880

Cys Leu Glu Val Thr Val Gln Gln Leu Thr Ala Gly Tyr His Ala Gly
 885 890 895

80

Gln Gly Glu Leu Ile Asn Phe Ser Ser Cys Leu Gln Ile Asn Leu Leu
 900 905 910

Ser Glu His Lys Pro Arg Ala Ser Gly Thr Pro Cys Phe Glu Leu Arg
 915 920 925

<210> 109

<211> 351

<212> PRT

<213> Homo sapien

<400> 109

Met Glu Pro Thr Glu Pro Met Glu Pro Thr Glu Pro Met Glu Pro Thr
 1 5 10 15

Glu Pro Met Glu Pro Thr Glu Pro Met Glu Pro Thr Glu Pro Met Glu
 20 25 30

Pro Ala Arg Ser Ala His Arg Gly Gly Glu Ala Leu Leu Arg Glu Leu
 35 40 45

Glu Val Leu Val Gln Asp Val Val Arg Thr Ser Ser Trp Trp Glu Arg
 50 55 60

His Gly Val Asp Cys Ala Ile Leu Ala Leu Ser Leu Phe Ala Leu Pro
 65 70 75 80

Ala Gly Phe Leu Cys Leu Arg Trp Glu Asn Ala Leu Val Phe Ala Ser
 85 90 95

Gly Ile Thr Ile Leu Gly Val Cys His Tyr Thr Leu Thr Val Lys Gly
 100 105 110

Ser His Leu Ala Thr His Gly Ala Leu Thr Glu Ser Lys Arg Trp Ser
 115 120 125

Lys Ile Trp Leu Leu Phe Phe Val Glu Val Cys Thr Ala Phe Thr Ala
 130 135 140

Glu His Ala Thr His Gly His Val Lys Met His His Ala Tyr Thr Asn
 145 150 155 160

Val Val Gly Leu Gly Asp Ser Ser Thr Trp Arg Leu Pro Cys Leu Asn
 165 170 175

Arg Tyr Val Tyr Met Phe Leu Ala Pro Phe Leu Leu Pro Ile Ala Thr
 180 185 190

81

Pro Leu Val Ala Val Glu Arg Leu Arg Lys Val Glu Leu Gly Thr Ala
 195 200 205

Leu Arg Thr Leu Ala Leu Ile Ser Leu Gly Leu Tyr Ser His Tyr Trp
 210 215 220

Leu Leu Leu Asn Val Ser Gly Phe Lys Asn Pro Ser Ser Ala Leu Gly
 225 230 235 240

Cys Met Phe Leu Thr Arg Ser Leu Leu Ala His Pro Tyr Leu His Val
 245 250 255

Asn Ile Phe Gln His Ile Gly Leu Pro Met Phe Ser Arg Asp Asn Lys
 260 265 270

Pro Arg Arg Ile His Met Met Ser Leu Gly Val Leu Asn Leu Ala Arg
 275 280 285

Leu Pro Val Leu Asp Trp Ala Phe Gly His Ser Ile Ile Ser Cys His
 290 295 300

Val Glu His His Leu Phe Pro Arg Leu Ser Asp Asn Met Cys Leu Lys
 305 310 315 320

Val Val Glu Gly Trp Ala Gly Gly Ala Gly Ile Lys Gly Leu Leu Glu
 325 330 335

Asp Gly Lys Glu Asp Ser Tyr Gly Leu Gly Ala Leu Leu Thr Leu
 340 345 350

<210> 110

<211> 715

<212> PRT

<213> Homo sapien

<400> 110

Met Arg Gln Ser Leu Leu Phe Leu Thr Ser Val Val Pro Phe Val Leu
 1 5 10 15

Ala Pro Arg Pro Pro Asp Asp Pro Gly Phe Gly Pro His Gln Arg Leu
 20 25 30

Glu Lys Leu Asp Ser Leu Leu Ser Asp Tyr Asp Ile Leu Ser Leu Ser
 35 40 45

82

Asn Ile Gln Gln His Ser Val Arg Lys Arg Asp Leu Gln Thr Ser Thr
 50 55 60

His Val Glu Thr Leu Leu Thr Phe Ser Ala Leu Lys Arg His Phe Lys
 65 70 75 80

Leu Tyr Leu Thr Ser Ser Thr Glu Arg Phe Ser Gln Asn Phe Lys Val
 85 90 95

Val Val Val Asp Gly Lys Asn Glu Ser Glu Tyr Thr Val Lys Trp Gln
 100 105 110

Asp Phe Phe Thr Gly His Val Val Gly Glu Pro Asp Ser Arg Val Leu
 115 120 125

Ala His Ile Arg Asp Asp Asp Val Ile Ile Arg Ile Asn Thr Asp Gly
 130 135 140

Ala Glu Tyr Asn Ile Glu Pro Leu Trp Arg Phe Val Asn Asp Thr Lys
 145 150 155 160

Asp Lys Arg Met Leu Val Tyr Lys Ser Glu Asp Ile Lys Asn Val Ser
 165 170 175

Arg Leu Gln Ser Pro Lys Val Cys Gly Tyr Leu Lys Val Asp Asn Glu
 180 185 190

Glu Leu Leu Pro Lys Gly Leu Val Asp Arg Glu Pro Pro Glu Glu Leu
 195 200 205

Val His Arg Val Lys Arg Arg Ala Asp Pro Asp Pro Met Lys Asn Thr
 210 215 220

Cys Lys Leu Leu Val Val Ala Asp His Arg Phe Tyr Arg Tyr Met Gly
 225 230 235 240

Arg Gly Glu Glu Ser Thr Thr Thr Asn Tyr Leu Ile Glu Leu Ile Asp
 245 250 255

Arg Val Asp Asp Ile Tyr Arg Asn Thr Ser Trp Asp Asn Ala Gly Phe
 260 265 270

Lys Gly Tyr Gly Ile Gln Ile Glu Gln Ile Arg Ile Leu Lys Ser Pro
 275 280 285

Gln Glu Val Lys Pro Gly Glu Lys His Tyr Asn Met Ala Lys Ser Tyr

83

| | | | | |
|---|-----|-----|-----|---------|
| 290 | | 295 | | 300 |
| Pro Asn Glu Glu Lys Asp Ala Trp Asp Val Lys Met Leu Leu Glu Gln | | | | |
| 305 | | 310 | | 315 320 |
| Phe Ser Phe Asp Ile Ala Glu Glu Ala Ser Lys Val Cys Leu Ala His | | | | |
| | 325 | | 330 | 335 |
| Leu Phe Thr Tyr Gln Asp Phe Asp Met Gly Thr Leu Gly Leu Ala Tyr | | | | |
| | 340 | | 345 | 350 |
| Val Gly Ser Pro Arg Ala Asn Ser His Gly Gly Val Cys Pro Lys Ala | | | | |
| | 355 | | 360 | 365 |
| Tyr Tyr Ser Pro Val Gly Lys Lys Asn Ile Tyr Leu Asn Ser Gly Leu | | | | |
| | 370 | | 375 | 380 |
| Thr Ser Thr Lys Asn Tyr Gly Lys Thr Ile Leu Thr Lys Glu Ala Asp | | | | |
| 385 | | 390 | | 395 400 |
| Leu Val Thr Thr His Glu Leu Gly His Asn Phe Gly Ala Glu His Asp | | | | |
| | 405 | | 410 | 415 |
| Pro Asp Gly Leu Ala Glu Cys Ala Pro Asn Glu Asp Gln Gly Gly Lys | | | | |
| | 420 | | 425 | 430 |
| Tyr Val Met Tyr Pro Ile Ala Val Ser Gly Asp His Glu Asn Asn Lys | | | | |
| | 435 | | 440 | 445 |
| Met Phe Ser Asn Cys Ser Lys Gln Ser Ile Tyr Lys Thr Ile Glu Ser | | | | |
| 450 | | 455 | | 460 |
| Lys Ala Gln Glu Cys Phe Gln Glu Arg Ser Asn Lys Val Cys Gly Asn | | | | |
| 465 | | 470 | | 475 480 |
| Ser Arg Val Asp Glu Gly Glu Glu Cys Asp Pro Gly Ile Met Tyr Leu | | | | |
| | 485 | | 490 | 495 |
| Asn Asn Asp Thr Cys Cys Asn Ser Asp Cys Thr Leu Lys Glu Gly Val | | | | |
| | 500 | | 505 | 510 |
| Gln Cys Ser Asp Arg Asn Ser Pro Cys Cys Lys Asn Cys Gln Phe Glu | | | | |
| | 515 | | 520 | 525 |
| Thr Ala Gln Lys Lys Cys Gln Glu Ala Ile Asn Ala Thr Cys Lys Gly | | | | |
| 530 | | 535 | | 540 |

84

Val Ser Tyr Cys Thr Gly Asn Ser Ser Glu Cys Pro Pro Pro Gly Asn
 545 550 555 560

Ala Glu Asp Asp Thr Val Cys Leu Asp Leu Gly Lys Cys Lys Asp Gly
 565 570 575

Lys Cys Ile Pro Phe Cys Glu Arg Glu Gln Gln Leu Glu Ser Cys Ala
 580 585 590

Cys Asn Glu Thr Asp Asn Ser Cys Lys Val Cys Cys Arg Asp Leu Ser
 595 600 605

Gly Arg Cys Val Pro Tyr Val Asp Ala Glu Gln Lys Asn Leu Phe Leu
 610 615 620

Arg Lys Gly Lys Pro Cys Thr Val Gly Phe Cys Asp Met Asn Gly Lys
 625 630 635 640

Cys Glu Lys Arg Val Gln Asp Val Ile Glu Arg Phe Trp Asp Phe Ile
 645 650 655

Asp Gln Leu Ser Ile Asn Thr Phe Gly Lys Phe Leu Ala Asp Asn Ile
 660 665 670

Val Gly Ser Val Leu Val Phe Ser Leu Ile Phe Trp Ile Pro Phe Ser
 675 680 685

Ile Leu Val His Cys Val Ile Arg Asn Trp Ile Asn Ser Met Asn Leu
 690 695 700

Cys Leu Cys Phe Thr Pro Val Thr Ser Lys Cys
 705 710 715

<210> 111

<211> 65

<212> PRT

<213> Homo sapien

<400> 111

Met Ala Ser Gly Gly Val Leu His Val Met Ser Gly Arg Lys Ser Gly
 1 5 10 15

Val Phe Leu Arg Gln Cys Val Phe Met Trp Ser Lys Gln Ser Lys Pro
 20 25 30

85

Val Ser Glu Ser Asn Pro Ser Met Thr Met Phe Pro His Leu Cys His
 35 40 45

Thr Leu Cys Glu Glu Leu Cys Pro His Phe Ser Leu Phe Asn Asn Leu
 50 55 60

Met
 65

<210> 112
 <211> 22
 <212> PRT
 <213> Homo sapien

<400> 112

Met Asp His Tyr Ile Tyr Pro Val Asn Phe Pro Gly Ser Asn Cys Gly
 1 5 10 15

Tyr Pro Asn Val Phe Glu
 20

<210> 113
 <211> 21
 <212> PRT
 <213> Homo sapien

<400> 113

Met Lys Phe Gln Leu Phe Ser Met His Lys Asn Arg Tyr Tyr Asp Ile
 1 5 10 15

Val His Tyr Thr Met
 20

<210> 114
 <211> 41
 <212> PRT
 <213> Homo sapien

<400> 114

Met Ala Lys Leu Val Thr Thr Ala Arg Ser Gly Ser Glu Arg Asp Asp
 1 5 10 15

Lys Glu Gly Glu Phe Lys Glu Pro Gln Thr Pro Gly Ile Phe Cys Ala
 20 25 30

Arg Ala Asn Asp Thr Glu Ser Ile Pro
 35 40

86

<210> 115
 <211> 42
 <212> PRT
 <213> Homo sapien

 <220>
 <221> MISC_FEATURE
 <222> (42)..(42)
 <223> x=any amino acid

<400> 115

Met Glu Lys Thr Leu Ala Trp Leu Ser Lys Asp Met Gly Ala Asn Ser
 1 5 10 15

Arg Leu Ala Leu Pro Ile Thr Tyr Cys Ala Gly Leu Thr Gln Ser Leu
 20 25 30

Pro Leu Thr Arg Ser Gln Phe Leu His Xaa
 35 40

<210> 116
 <211> 28
 <212> PRT
 <213> Homo sapien

<400> 116

Met Glu Thr His Leu Leu Met Arg Lys Gln Phe Thr Thr Cys Ser Ile
 1 5 10 15

Glu His Ser Tyr Leu Glu Phe Asn Thr His Leu Tyr
 20 25

<210> 117
 <211> 97
 <212> PRT
 <213> Homo sapien

<400> 117

Met Pro Leu Ala Val Thr Gly Thr Cys His Ala Cys Ser Phe Ile Gly
 1 5 10 15

His Cys Thr Cys Leu Leu Phe Ala Phe Lys Ala Leu Pro Leu Asp Ile
 20 25 30

Arg Ile Ala Ser Phe Phe Ala Ser Phe Arg Phe Leu Thr Ile Cys His
 35 40 45

Leu Leu Gly Glu Ala Phe Tyr Asp His Leu Thr Cys Asn Ser Ser Ser
 50 55 60

87

Thr Pro Tyr Tyr Leu Ser Pro Ser Ser Val Leu Phe Phe Phe Thr Val
65 70 75 80

Tyr Phe Trp Leu Ile Val Ser Leu Pro Gln Asp Ile Val Ile Ser Gly
85 90 95

Gly

<210> 118

<211> 137

<212> PRT

<213> Homo sapien

<400> 118

Met Cys Leu Gly Ser Gly Ile Thr Trp Leu Gly Pro Gln Ile Phe Ser
1 5 10 15

Ser Ala Trp Lys Arg Phe Thr Ser Ser Ala Ser Ser Arg Cys Gly Ser
20 25 30

Arg Gly Ile Asp Gly Leu Leu Thr Ser Thr Phe Ser Phe Pro Ala His
35 40 45

Leu Ala Leu Leu Leu Gly His Val Ser Pro Val His Leu Gln Glu Thr
50 55 60

Ser Val Asp Ala Pro Cys Leu Leu Thr Leu Ser Pro Ala His Thr Glu
65 70 75 80

Leu Val Leu Arg Gly Asn Leu Cys Leu Cys Cys Cys Leu Cys Leu Glu
85 90 95

Arg Pro Cys Pro Thr Pro Ser His Ser Cys Leu Ser Val Ile Phe Pro
100 105 110

Met Ser Val Leu Arg Glu Pro Phe Leu Ala Thr Pro Ser Lys Gly Val
115 120 125

Leu Gly Gln Val Trp Trp Pro Thr Pro
130 135

<210> 119

<211> 51

<212> PRT

<213> Homo sapien

88

<400> 119

Met Ser Val Leu Arg Glu Pro Phe Leu Ala Thr Pro Ser Lys Gly Val
 1 5 10 15

Leu Gly Gln Val Trp Trp Pro Thr Pro Val Ile Ser Glu Leu Trp Glu
 20 25 30

Thr Glu Val Gly Arg Ser Leu Glu Ala Arg Gly Ser Arg Pro Thr Trp
 35 40 45

Ala Thr Tyr
 50

<210> 120

<211> 89

<212> PRT

<213> Homo sapien

<400> 120

Met Ser Leu Cys Leu Ala Cys Thr Leu Cys Leu Gly Cys Ser Leu His
 1 5 10 15

Pro Ala Leu Pro Gly Trp Ala Ser Asp Ser Asn Leu Leu Gly Leu Pro
 20 25 30

Pro His Leu Cys Asp Ser Gly Ile Ile Pro Lys Ala Val Val Arg Ile
 35 40 45

Leu Gln Glu Asn Ala Trp Lys Val Leu Gly Thr Met Leu Ser Pro Tyr
 50 55 60

Asp Thr His Ser Cys Val Leu Leu Ser Leu Thr Tyr Cys Phe Ser Thr
 65 70 75 80

Thr Thr Thr Ile Arg Ile Leu Lys Val
 85

<210> 121

<211> 64

<212> PRT

<213> Homo sapien

<400> 121

Met Pro Ala Val Thr Ile Thr Ile Met Tyr Phe Cys Cys Cys His Thr
 1 5 10 15

89

Lys Met Asn Asn Asn Ile Leu Ser His Leu Lys Pro Lys Arg Arg Asn
 20 25 30

Gln Trp Glu Gly Cys Leu Gln Pro Ala His Gln His Arg Lys Gly Ser
 35 40 45

Pro Ala Ser Tyr Pro Asn Ser Gln Arg Pro Asn Pro Arg Leu Leu His
 50 55 60

<210> 122
 <211> 34
 <212> PRT
 <213> Homo sapien

<400> 122

Met Arg Glu His Asn Asn His Asn Asp Asn Glu Cys Ser Val Val Lys
 1 5 10 15

Leu Thr Gly Thr Leu Leu Phe Leu Leu Ser Val Gln Pro Asn Ala Ser
 20 25 30

Ala Asp

<210> 123
 <211> 151
 <212> PRT
 <213> Homo sapien

<400> 123

Met Tyr Gly Cys Tyr Thr Pro Thr Ala Tyr Ser Thr Arg Ser Ala Pro
 1 5 10 15

Glu Glu Asp Trp Val Lys Leu Cys Lys Phe Gly Phe Pro Gly Asn Ala
 20 25 30

Leu His Tyr Ser Ala Pro Asp Leu Pro Thr Thr Pro Val Gly Thr Arg
 35 40 45

Ser Ser Thr His Leu Ala Glu Leu Met Thr Ala Trp Ala Gln Arg Ser
 50 55 60

Ala His Cys Ala Asn Thr Arg Thr Gly Ile Ala Pro Leu Pro Glu Pro
 65 70 75 80

Pro His Arg Ala Pro Phe Lys Glu Leu Ala Thr Pro Leu Thr Cys Lys
 85 90 95

90

Gln Pro Pro Thr Leu Lys Leu Ile Arg Thr Arg Val Phe His Pro Lys
 100 105 110

Gly Leu Cys Cys Gly Arg Cys Ser Asp Pro Arg Arg Gly Arg Glu Val
 115 120 125

Pro Lys Ala Thr Ala Arg Gly Trp Gly Thr Pro Leu Leu Thr Leu Val
 130 135 140

Leu Asp Phe Glu Gly Pro Asn
 145 150

<210> 124
 <211> 59
 <212> PRT
 <213> Homo sapien

<400> 124

Met Asn Cys Leu Trp Ile Leu Leu Ser Ile Ser Leu Val Pro Phe Leu
 1 5 10 15

Gln Leu Tyr Gly Thr Leu Ser Ser Cys Thr Pro Glu Ala Pro Gln Leu
 20 25 30

Gly Lys Val Ser Gln Arg Tyr Gln Glu Tyr Met Leu Arg Gly His Phe
 35 40 45

Lys Val Phe His Arg Arg Leu Cys Leu Gly Lys
 50 55

<210> 125
 <211> 83
 <212> PRT
 <213> Homo sapien

<400> 125

Met Asp Leu Cys Ile Ser Arg Ser Cys Ser Leu Leu Ser Val Gly Val
 1 5 10 15

Leu Thr Arg Lys Ser Leu Ala Val Gln His Arg Cys Thr Ala Thr Leu
 20 25 30

Gly Ala Val Glu Trp Ser Ala Gly Ser Gln Ser Asn Cys Phe Pro Leu
 35 40 45

Asp Leu Gly Leu Val Leu Phe Ser Lys Tyr Gln Ser Tyr Leu Lys Met
 50 55 60

91

Lys Ala Cys Lys Pro Leu Leu Lys Ser Ile Glu Ser Glu Arg Glu Arg
 65 70 75 80

Leu Phe Cys

<210> 126
 <211> 34
 <212> PRT
 <213> Homo sapien

<400> 126

Met Lys Lys Ile Phe Lys Phe Asn Lys Ile Ile Ile Tyr Cys Pro Lys
 1 5 10 15

Ile Ile Ile Lys Ser Leu Leu Val Glu Leu Ser Gln Asn Lys Gly Asn
 20 25 30

Ser Phe

<210> 127
 <211> 862
 <212> PRT
 <213> Homo sapien

<400> 127

Met Met Pro Trp Ala Leu Gln Lys Lys Arg Glu Ile His Met Ala Lys
 1 5 10 15

Ala His Arg Arg Arg Ala Ala Arg Ser Ala Leu Pro Met Arg Leu Thr
 20 25 30

Ser Cys Ile Phe Arg Arg Pro Val Thr Arg Ile Arg Ser His Pro Asp
 35 40 45

Asn Gln Val Arg Arg Arg Lys Gly Asp Glu His Leu Glu Lys Pro Gln
 50 55 60

Gln Leu Cys Ala Tyr Arg Arg Leu Gln Ala Leu Gln Pro Cys Ser Ser
 65 70 75 80

Gln Gly Glu Gly Ser Ser Pro Leu His Leu Glu Ser Val Leu Ser Ile
 85 90 95

Leu Ala Pro Gly Thr Ala Ser Glu Ser Leu Asp Arg Ala Gly Ala Glu

92

| | | |
|---|-----|-----|
| 100 | 105 | 110 |
| Arg Val Arg Ser Pro Leu Glu Pro Thr Pro Gly Arg Phe Pro Ala Val | | |
| 115 | 120 | 125 |
| Ala Gly Gly Pro Thr Pro Gly Met Gly Cys Gln Leu Pro Pro Pro Leu | | |
| 130 | 135 | 140 |
| Ser Gly Gln Leu Val Thr Pro Ala Asp Ile Arg Arg Gln Ala Arg Arg | | |
| 145 | 150 | 155 |
| Val Lys Lys Ala Arg Glu Arg Leu Ala Lys Ala Leu Gln Ala Asp Arg | | |
| 165 | 170 | 175 |
| Leu Ala Arg Arg Ala Glu Met Leu Thr Gly Arg Gln Thr Ile Pro Lys | | |
| 180 | 185 | 190 |
| Tyr Asn Cys Gln Thr Gln Asn Lys Thr Val Ala Val Met Pro Gly Thr | | |
| 195 | 200 | 205 |
| Thr Ser Val Ser Ser Thr Lys Ile Gly Ala Leu Leu Gly Cys Tyr Arg | | |
| 210 | 215 | 220 |
| Ser Pro Arg Gly Ala Val Leu Arg Ala Leu Arg Asp Lys Gly Gly Ala | | |
| 225 | 230 | 235 |
| Ser Cys Phe Pro Ala Ala Leu Leu Leu Ser Glu Ser Gln Asp Gly Gly | | |
| 245 | 250 | 255 |
| Cys Arg Pro Ala Pro Asn Cys Arg Ala Met Leu Trp Phe Pro Trp Ile | | |
| 260 | 265 | 270 |
| Ala Val Gly Trp Cys Leu Leu Gly Val Glu Val Asp Ala Leu Trp Val | | |
| 275 | 280 | 285 |
| Thr Val Tyr Ala Trp Gly Pro Gly Arg Ser Gln Ala Ala Ser Gln Arg | | |
| 290 | 295 | 300 |
| Glu Gly His Thr Glu Gly Gly Glu Leu Val Asn Glu Leu Leu Lys Ser | | |
| 305 | 310 | 315 |
| Trp Leu Lys Gly Leu Val Thr Phe Glu Asp Val Ala Val Glu Phe Thr | | |
| 325 | 330 | 335 |
| Gln Glu Glu Trp Ala Leu Leu Asp Pro Ala Gln Arg Thr Leu Tyr Arg | | |
| 340 | 345 | 350 |

Asp Val Met Leu Glu Asn Cys Arg Asn Leu Ala Ser Leu Gly Asn Gln
 355 360 365

Val Asp Lys Pro Arg Leu Ile Ser Gln Leu Glu Gln Glu Asp Lys Val
 370 375 380

Met Thr Glu Glu Arg Gly Ile Leu Ser Gly Thr Cys Pro Asp Val Glu
 385 390 395 400

Asn Pro Phe Lys Ala Lys Gly Leu Thr Pro Lys Leu His Val Phe Arg
 405 410 415

Lys Glu Gln Ser Arg Asn Met Lys Met Glu Arg Asn His Leu Gly Ala
 420 425 430

Thr Leu Asn Glu Cys Asn Gln Cys Phe Lys Val Phe Ser Thr Lys Ser
 435 440 445

Ser Leu Thr Arg His Arg Lys Ile His Thr Gly Glu Arg Pro Tyr Gly
 450 455 460

Cys Ser Glu Cys Gly Lys Ser Tyr Ser Ser Arg Ser Tyr Leu Ala Val
 465 470 475 480

His Lys Arg Ile His Asn Gly Glu Lys Pro Tyr Glu Cys Asn Asp Cys
 485 490 495

Gly Lys Thr Phe Ser Ser Arg Ser Tyr Leu Thr Val His Lys Arg Ile
 500 505 510

His Asn Gly Glu Lys Pro Tyr Glu Cys Ser Asp Cys Gly Lys Thr Phe
 515 520 525

Ser Asn Ser Ser Tyr Leu Arg Pro His Leu Arg Ile His Thr Gly Glu
 530 535 540

Lys Pro Tyr Lys Cys Asn Gln Cys Phe Arg Glu Phe Arg Thr Gln Ser
 545 550 555 560

Ile Phe Thr Arg His Lys Arg Val His Thr Gly Glu Gly His Tyr Val
 565 570 575

Cys Asn Gln Cys Gly Lys Ala Phe Gly Thr Arg Ser Ser Leu Ser Ser
 580 585 590

His Tyr Ser Ile His Thr Gly Glu Tyr Pro Tyr Glu Cys His Asp Cys
 595 600 605

Gly Arg Thr Phe Arg Arg Arg Ser Asn Leu Thr Gln His Ile Arg Thr
 610 615 620

His Thr Gly Glu Lys Pro Tyr Thr Cys Asn Glu Cys Gly Lys Ser Phe
 625 630 635 640

Thr Asn Ser Phe Ser Leu Thr Ile His Arg Arg Ile His Asn Gly Glu
 645 650 655

Lys Ser Tyr Glu Cys Ser Asp Cys Gly Lys Ser Phe Asn Val Leu Ser
 660 665 670

Ser Val Lys Lys His Met Arg Thr His Thr Gly Lys Lys Pro Tyr Glu
 675 680 685

Cys Asn Tyr Cys Gly Lys Ser Phe Thr Thr Ser Thr Thr Ser Thr Thr
 690 695 700

Ala Ser Ala Thr Ser Thr Thr Ala Ser Trp Ile His His Ser Asn His
 705 710 715 720

His Ile His Tyr Ser Ile His Tyr Arg Ile His Tyr Ser Asp Ser Asn
 725 730 735

Thr Val Ala Ser His Ser Trp Gly Tyr Ala Ser Gly Asn Gly Arg Val
 740 745 750

Glu Glu Gly Gly Arg Asp Pro Arg Val Val Ala Ser Thr Gly Asn Asp
 755 760 765

Pro Glu Ala Ser Asp Asn Asp Glu Asp Asn Asp Gly Asp Asp Lys Asn
 770 775 780

Gly His Asp Gly Asp Asp Ser Asp Gly Asn Glu Gly Asp Gly Asp Glu
 785 790 795 800

Met Val Met Met Met Met Ile Leu Thr Met Met Val Met Thr Val Ile
 805 810 815

Met Met Met Val Met Leu Met Thr Val Met Glu Met Ala Asn Leu Thr
 820 825 830

95

Thr Pro Tyr Cys Gly Thr Leu Gly Gln Ser Leu Glu Glu Cys Glu Glu
 835 840 845

Phe Arg Arg Asn Phe His Ala Ala Ser Gly Lys Leu Pro Gly
 850 855 860

<210> 128

<211> 187

<212> PRT

<213> Homo sapien

<400> 128

Met Met Pro Trp Ala Leu Gln Lys Lys Arg Glu Ile His Met Ala Lys
 1 5 10 15

Ala His Arg Arg Arg Ala Ala Arg Ser Ala Leu Pro Met Arg Leu Thr
 20 25 30

Ser Cys Ile Phe Arg Arg Pro Val Thr Arg Ile Arg Ser His Pro Asp
 35 40 45

Asn Gln Val Arg Arg Arg Lys Gly Asp Glu His Leu Glu Lys Pro Gln
 50 55 60

Gln Leu Cys Ala Tyr Arg Arg Leu Gln Ala Leu Gln Pro Cys Ser Ser
 65 70 75 80

Gln Gly Glu Gly Ser Ser Pro Leu His Leu Glu Ser Val Leu Ser Ile
 85 90 95

Leu Ala Pro Gly Thr Ala Gly Glu Ser Leu Asp Arg Ala Gly Ala Glu
 100 105 110

Arg Val Arg Ser Pro Leu Glu Pro Thr Pro Gly Arg Phe Pro Ala Val
 115 120 125

Ala Gly Gly Pro Thr Pro Gly Met Gly Cys Gln Leu Pro Pro Pro Leu
 130 135 140

Ser Gly Gln Leu Val Thr Pro Ala Asp Ile Arg Arg Gln Ala Arg Arg
 145 150 155 160

Val Lys Lys Ala Arg Glu Arg Leu Ala Lys Ala Leu Gln Ala Asp Arg
 165 170 175

Leu Ala Arg Gln Ala Glu Met Leu Thr Cys Arg
 180 185

96

<210> 129
 <211> 11
 <212> PRT
 <213> Homo sapien

<400> 129

Met Arg Lys Leu His Phe Glu Lys Tyr Arg Tyr
 1 5 10

<210> 130
 <211> 109
 <212> PRT
 <213> Homo sapien

<400> 130

Met Ala Gly Asn Leu Thr Gln Pro Ala Gly Gln His Thr Thr Pro Glu
 1 5 10 15

Pro Ser His Ala Gly Gly Cys Pro Gln Val Pro Lys Gly Leu Glu Gln
 20 25 30

Ala Thr Leu Gly Asp Cys Leu His Pro Ile Leu Gln Arg Ala Val Cys
 35 40 45

Gln Arg Val Pro Ala Ala Ser Gln Thr Ala Ile Ile His Gly Glu Met
 50 55 60

Leu Ala Thr Val Pro Ser Thr Lys Val Tyr Ser Gln Cys Ile Ser Leu
 65 70 75 80

Arg Leu Tyr Gly Gln Arg Val Gly Tyr His Leu Lys Gly Gln Lys Ala
 85 90 95

Glu Pro Leu Ser Cys Trp Gly Ser Gly His Gln Leu Val
 100 105

<210> 131
 <211> 127
 <212> PRT
 <213> Homo sapien

<400> 131

Met Leu Ile Ile Leu Ala Leu Trp Asp Ala Glu Val Glu Gly Pro Leu
 1 5 10 15

Glu Ala Ser Leu Gln Ala Pro Glu Glu Thr Ser Ser Met Ile Pro Cys
 20 25 30

97

Leu Arg Gly Arg Phe Leu Arg Thr Leu Pro Trp Gly Leu Lys Val Ala
 35 40 45

Thr Pro Pro Gly Thr Tyr Ala Pro Gly Ala Glu Phe Gly Leu Pro Lys
 50 55 60

Ser Val Cys Val Arg Leu Cys Leu Cys Ala Cys Leu His Val Cys Val
 65 70 75 80

Phe Ala Cys Met Ser Val Cys Leu Cys Gly Leu Cys Val Cys Val Tyr
 85 90 95

Thr Ser Val Cys Leu Ser Leu Cys Ile Phe Ala Cys Val Ser Met Cys
 100 105 110

Leu Cys Ala Cys Leu Cys Val Tyr Met Ser Val Ser Leu Tyr Leu
 115 120 125

<210> 132
 <211> 78
 <212> PRT
 <213> Homo sapien

<400> 132

Met His Ile Tyr Val Ile Ser Thr Asn Ser Pro Phe Val Lys Thr Asn
 1 5 10 15

His Met Ala Lys Pro Ile Val Arg Gly Lys Tyr Thr Val Phe Phe Arg
 20 25 30

Ala Asn Tyr Arg Phe Thr Gln Gln Lys Ala Trp Ile Gln Glu Arg Val
 35 40 45

Lys Gly Ser Ile Ile His Ser Thr Asn Arg Arg Ile Ser Arg Leu Arg
 50 55 60

Pro Lys Asn Arg Cys Gln Gly His Val His His Thr Asp Ile
 65 70 75

<210> 133
 <211> 56
 <212> PRT
 <213> Homo sapien

<400> 133

Met Gly Gly Arg Glu Asp Arg Pro Gly Val Trp Asp Val Thr Ser Ala

```

1              5              10              15
His Thr Gln Gly Arg Pro Ala Gln Gly Trp Gly Leu Leu Leu Cys Leu
          20              25              30
Met Cys Ala Gly Ser Asp Thr Gly Leu Val Trp Gly Arg Leu Arg Thr
          35              40              45
Val Lys Met Lys Asn Lys Arg Lys
          50              55

```

<400> 134

Ile Gly Ser Gln Gly
50

<400> 135

Met Tyr Leu Arg Val Leu Thr Leu Leu Ser Val Thr Met Phe Tyr
35 40 45

```
<210> 136
<211> 162
<212> PRT
<213> Homo sapien
```

99

<400> 136

Met Val Arg Arg Phe Leu Val Thr Leu Arg Ile Arg Arg Ala Cys Gly
 1 5 10 15

Pro Pro Arg Val Arg Val Phe Val Val His Ile Pro Arg Leu Thr Gly
 20 25 30

Glu Trp Ala Ala Pro Gly Ala Pro Ala Ala Val Ala Leu Val Leu Met
 35 40 45

Leu Leu Arg Ser Gln Arg Leu Gly Gln Gln Pro Leu Pro Arg Arg Pro
 50 55 60

Gly Arg Lys Gly Pro Arg Lys Val Arg Gly Ala Phe Gly Thr Cys Phe
 65 70 75 80

Val Cys Thr Val Glu Pro Gly Pro Pro Phe Thr Gln Pro Leu Asn Arg
 85 90 95

Gly Lys Leu Arg Arg Thr Glu Leu Leu Asn Pro Pro Gln Lys Ser Glu
 100 105 110

Ser Gly Gly Pro Val Arg Tyr Gly Ala Arg Arg Gln Arg Leu Pro Glu
 115 120 125

Met Leu Thr Arg Ile Val Gly Ser Val Thr Ala Ala Arg Ser Lys His
 130 135 140

Leu Leu Leu Gly Gly Val Leu Gln Pro Thr Gly Lys Trp Ala Phe Phe
 145 150 155 160

Val Ser

<210> 137

<211> 46

<212> PRT

<213> Homo sapien

<400> 137

Met Arg Ile Cys Leu Cys Phe His Leu Leu Ser Leu Cys Leu Cys Leu
 1 5 10 15

Ser Asn Leu Pro Pro Gln Glu His Gln Gly Asn Gln Cys Cys Arg Pro
 20 25 30

100

Arg Leu Asn Leu Arg Phe Leu Asn Gly Arg Asn Glu Ser Thr
 35 40 45

<210> 138
 <211> 47
 <212> PRT
 <213> Homo sapien

<400> 138

Met Leu Phe Gln Gln Leu Leu Lys Leu Phe Cys Gln Met Leu Phe Leu
 1 5 10 15

Gln Cys Glu Gln Asn Ile Asn Arg Ile Lys Asn Ile Asn Arg Thr Glu
 20 25 30

Asn Lys Asn His Val Ile Ile Ser Ile Leu Ile Gly Leu His Met
 35 40 45

<210> 139
 <211> 68
 <212> PRT
 <213> Homo sapien

<400> 139

Met Leu Ala Gly Ser Leu Leu Leu Ala Phe Glu Arg Gly Ala Ala Val
 1 5 10 15

Ser Ser Val Cys Arg Val Lys Lys Arg Gly Gln Ser Ser Glu Glu Met
 20 25 30

Asn Pro Ser Arg Lys Gly Cys Leu Thr Ser Arg Arg Glu Asp Asn Gln
 35 40 45

Glu Gly Leu Trp Val Ser Leu Leu Ser Tyr Thr Ser Phe Gln Val Thr
 50 55 60

Trp Asp His Gly
 65

<210> 140
 <211> 13
 <212> PRT
 <213> Homo sapien

<400> 140

Met Trp Gln Val Glu Asp Gly Phe Leu Ser Ser Leu Thr
 1 5 10

101

<210> 141
 <211> 108
 <212> PRT
 <213> Homo sapien

<400> 141

Met Gln Pro Arg Leu Phe Ser Cys Ala Tyr Val Ser Arg Met Pro Thr
 1 5 10 15

Phe Leu Phe Asn Ala Met Gln Ile Ser Lys Lys Ser Leu Thr Met Lys
 20 25 30

Ala Ala Cys Arg Gln Lys Ser Phe Gln Cys Ile His Val His Ile His
 35 40 45

Thr Leu His Arg His Met Ser Ser Tyr Lys Glu Ile Thr Gln Asp Gln
 50 55 60

Asn Phe Thr Tyr Cys Gln Phe Leu Lys Val Gln Pro His Phe Gln Lys
 65 70 75 80

Pro Gln Thr Ser Ile Arg Ser Asn Glu Ile Tyr Ile Phe Leu Ile Leu
 85 90 95

Gly Lys Cys Asn Ile Pro Val Leu Glu Gln Gly Glu
 100 105

<210> 142
 <211> 38
 <212> PRT
 <213> Homo sapien

<400> 142

Met Ser Leu Leu Ile Leu Thr Met Lys Leu Lys Lys Lys Lys His Leu
 1 5 10 15

Lys Ser Thr Glu Ile Arg Val Lys Val Thr Val Ile Thr Tyr Leu Tyr
 20 25 30

His Asn Ile Leu Val Asp
 35

<210> 143
 <211> 74
 <212> PRT
 <213> Homo sapien

<400> 143

102

Met His Thr His His Tyr Tyr Gly Val Ser Tyr Pro Gln Leu Ser Pro
 1 5 10 15

Asp Gln Ala Leu Lys Ala Gly Arg Ala Arg Ser Gly Ile Pro Gly Lys
 20 25 30

Gly Trp Glu Gly Leu Ala Leu Arg Lys Gly Cys Glu Thr Gly Met Arg
 35 40 45

Trp Gly Glu Cys Ser Glu Gly Gly Gly Lys Gly Val Pro Ala Gly Gly
 50 55 60

Val Cys Ser Ser Thr Ala Glu Ala Ala Glu
 65 70

<210> 144

<211> 119

<212> PRT

<213> Homo sapien

<400> 144

Met Ile Ala Arg Gly Leu Ala Cys Cys Leu Leu Asp Ser Phe Leu Leu
 1 5 10 15

Leu Phe Ser Leu Pro Val Gly Trp Thr Cys His Cys Cys Thr Cys Ala
 20 25 30

Phe Ala Phe Ser Tyr Ser Phe Phe His Leu Leu Leu Ser Ile Cys Asp
 35 40 45

Thr Ser Trp Cys Val Ser Tyr Arg Trp Pro Ser Ser Cys Cys Arg Ser
 50 55 60

Leu Ala Leu Pro Gly Val Ser Ser Leu Ser Arg Val Pro Pro Leu Leu
 65 70 75 80

Pro Ser Cys Arg Leu Arg Phe Gly Gly Pro Ser Val Arg Val Arg Phe
 85 90 95

Pro Ile Val Pro Gly Tyr Pro Met Trp Ala Pro Leu Ala Arg Ser Pro
 100 105 110

Pro Phe Gly Asn Arg Phe Arg
 115

<210> 145

103

<211> 23
 <212> PRT
 <213> Homo sapien

<400> 145

Met Ile Lys Gly Ile Gly Lys Ser Thr Lys Thr Lys Ser Ser Asp Glu
 1 5 10 15

Thr Gln Glu Ser Gly Arg Arg
 20

<210> 146
 <211> 33
 <212> PRT
 <213> Homo sapien

<400> 146

Met Leu Asn Leu Leu Ile Ile Ser Pro Leu Asp Cys Lys Ile Tyr Val
 1 5 10 15

Gly Arg Asp Lys Ile Val Ser Val Leu Ile Val Ser Pro Lys Pro Leu
 20 25 30

Glu

<210> 147
 <211> 130
 <212> PRT
 <213> Homo sapien

<400> 147

Met Gln Leu Gln Ala Pro Trp Pro Gln Cys Ser Ser Lys His Gln Val
 1 5 10 15

Cys Thr Cys Leu Gly Gln Ser Val Leu Gly Ile Pro Ser Ala Leu Asn
 20 25 30

Val Val Leu Pro Glu Lys Cys Ile Val His Ser Tyr Ile Leu Lys Val
 35 40 45

Ser Leu His Cys Tyr Leu Ser Arg Ser Ser Leu Asn Ile Tyr Ile Ser
 50 55 60

Ile Pro Phe Pro Pro Cys Phe Met Tyr Val His Ser Thr His His Asp
 65 70 75 80

Leu Thr Leu Leu Ser Ile Tyr Ser Phe Thr Val Cys Phe Pro Ile Pro

104
 85 90 95
 Lys Tyr Lys Leu Ser Lys Asp Arg Asn Phe Cys Ala Leu Leu Leu Asn
 100 105 110
 Leu Gln Phe Val Glu Lys Cys Leu Pro Tyr Ile Lys His Ser Val Asn
 115 120 125
 Val Tyr
 130

<210> 148
 <211> 51
 <212> PRT
 <213> Homo sapien

<400> 148

Met Ser Ser Lys Glu Val Ser Leu Asp Ser Leu Leu Leu Gly Arg Leu
 1 5 10 15

Met Arg Gly Tyr Gln Arg Thr Glu Phe Asn Cys Leu Trp Ile Leu Leu
 20 25 30

Lys Thr Leu Arg Ala Ser Gln Gly Ala Tyr Met Pro Arg Leu Ser Leu
 35 40 45

Gly Ile Gly
 50

<210> 149
 <211> 42
 <212> PRT
 <213> Homo sapien

<400> 149

Met Ala Ile Ser Leu Ile Val Cys Lys Ile Tyr Lys Ala Tyr Leu Asn
 1 5 10 15

Lys Phe Phe Leu Phe Ile Val Met Lys Thr Arg Ser Ser Leu Ser Gln
 20 25 30

His Thr Ser Ser Asn Tyr Leu Gln Pro Leu
 35 40

<210> 150
 <211> 44
 <212> PRT
 <213> Homo sapien

105

<400> 150

Met Ser Glu Arg Ala Pro Lys Thr Phe Ser Cys Phe Phe Val Phe Phe
1 5 10 15

Tyr Pro Thr Trp His Thr Trp His Gln Pro Asn Cys Phe Ile Thr Glu
20 25 30

Asn Lys Tyr Phe Leu Pro Lys Tyr Leu Phe His Arg
35 40

<210> 151

<211> 68

<212> PRT

<213> Homo sapien

<400> 151

Met Tyr Ile Phe Phe Gly Ser Ala Val Leu Leu Leu Arg Ile Phe His
1 5 10 15

Pro Thr Gly Tyr Lys Lys Met Phe Thr Ile Val Leu Phe Leu Leu Leu
20 25 30

Glu Ser Gly Thr Gln Tyr Lys Gln Gln Ser Leu Arg Asp Trp Ser His
35 40 45

Val Ser Thr Gln Glu His Arg Asp Gly Ile Gln Arg Leu Ile Ile Met
50 55 60

Trp Leu Ser Leu
65

<210> 152

<211> 91

<212> PRT

<213> Homo sapien

<400> 152

Met Trp Ser Ile His Ser Leu Lys Leu Thr Leu Ser Glu Thr Thr Asp
1 5 10 15

Ser Arg Ser Ser Asn Asn Ala Thr Ala Phe Asn Met Val Ser Leu Ile
20 25 30

Thr Leu His Asp His Lys Glu Ala Ala Ile Tyr Ser Phe Cys Leu Arg
35 40 45

106

Leu Tyr Leu Gln Glu Pro Ile Asp Lys Val Asn Cys Phe Tyr Phe Lys
 50 55 60

Leu Gln Ser Val Glu Tyr His His Ser Gln Ser Lys Leu His Glu Asn
 65 70 75 80

Arg Ala Tyr Ile Phe Val Thr His Glu Gly Gly
 85 90

<210> 153
 <211> 156
 <212> PRT
 <213> Homo sapien

<400> 153

Met His Arg Ser Gly His Leu Ser Trp Val Thr Ser Leu Val Ala Ile
 1 5 10 15

Gly Leu Gly Ser Gly Ala Ala Cys Thr Glu Arg Arg Pro Trp Ala Gln
 20 25 30

Arg Ala Ala Gly Ala Pro Gly Gly Ser Phe Thr Glu Ala Gly Leu Gly
 35 40 45

Leu Ala Pro Glu Pro Val Arg Ser Gly Val Gly Ser Gly Ala Gly Asp
 50 55 60

Ser Ala Met Ala Thr Ala Gly Ala Ala Cys Val Ser Ala Val Ala Pro
 65 70 75 80

Ser Asp Pro Ile Gly Ala Arg Asp Leu Ser Gly Ala Ala Ala Gly
 85 90 95

Pro Gly Arg Thr Gly Glu Gly Asp Ala Gly Arg His Gly Asp Leu Gly
 100 105 110

Arg Arg Ala Gly Gly Gly Gly His Ala Gly Gly Ser Ala Gly Arg Gly
 115 120 125

Arg Gly Gly His Arg Ala Ala Glu Lys Val Pro Met Ala Trp Gly Ser
 130 135 140

Gly Ala Arg Val Gly Lys Thr Gln Asp Ile Ala Ser
 145 150 155

<210> 154
 <211> 22

107

<212> PRT

<213> Homo sapien

<400> 154

Met Phe Arg Asp Leu Gly Tyr Phe Lys Glu Val Ser Ala Ala Val Val
1 5 10 15

Thr Asp Val Glu Leu Leu
20

<210> 155

<211> 46

<212> PRT

<213> Homo sapien

<400> 155

Met Gly Glu Glu Leu Gly Arg Gln Glu Thr Ala Tyr Asn Leu Thr Asp
1 5 10 15

Gln Glu Lys Gly Ser Gly Phe Tyr Ser Lys Cys Lys Gly Lys Leu Trp
20 25 30

Glu Arg Phe Lys Gln Ile Ser Lys Thr Ile Phe Tyr Leu Gln
35 40 45

<210> 156

<211> 51

<212> PRT

<213> Homo sapien

<400> 156

Met Tyr Thr Ser Asn Gln Tyr Ile Tyr Leu Lys Leu Trp Leu Glu Leu
1 5 10 15

Thr Ala Leu Met Asn Thr Met Cys Pro Gln Lys Leu Thr Ile Glu Gly
20 25 30

Phe Lys Thr Lys Lys Leu His Thr Thr Thr Phe Leu His Thr Leu Asn
35 40 45

Lys Lys Ile
50

<210> 157

<211> 72

<212> PRT

<213> Homo sapien

<400> 157

108

Met Ala Ser Leu Asn Ser Lys Ile Leu Pro Ser Ser Lys Leu Leu Leu
 1 5 10 15

Gly Leu Met Val Lys Ile Leu Thr Asn Gly Ser Lys Ile Pro Tyr Leu
 20 25 30

Pro His Pro Pro Ile Pro Thr Val Leu Arg Pro Leu Ser Pro Arg Ser
 35 40 45

Trp Ile Leu Ser Ser Gln Val Val Ser Ala Cys His Gln Lys Glu Cys
 50 55 60

Asn Ser Val Leu Asp Leu His Thr
 65 70

<210> 158
 <211> 32
 <212> PRT
 <213> Homo sapien

<400> 158

Met Lys Cys Arg Gln Met Ala Arg Cys Cys Glu Val Lys Trp Leu Trp
 1 5 10 15

His Leu Val Ala Thr Glu Thr Thr Val Ile Lys Ile Met Ile Ile Thr
 20 25 30

<210> 159
 <211> 19
 <212> PRT
 <213> Homo sapien

<400> 159

Met Ser Leu Val Ser Gly Glu His Lys Ile Ile Asp Ser Ala Glu Val
 1 5 10 15

Trp Ser Lys

<210> 160
 <211> 46
 <212> PRT
 <213> Homo sapien

<400> 160

Met Tyr Thr Glu Ile Ser Lys Val Leu Gln Lys Tyr Pro Lys Ser Arg
 1 5 10 15

109

Val Ser Ile Ser His Phe Thr Ser Ser Ala Cys Phe Thr Pro Ile Leu
 20 25 30

Asp Cys Phe Ile Ser Glu Leu Asp Val Ile Pro Arg Val Arg
 35 40 45

<210> 161

<211> 47

<212> PRT

<213> Homo sapien

<400> 161

Met Pro Gln Trp His Gln Pro Ser His Leu Glu Lys Lys Ser Leu Ser
 1 5 10 15

Ser Val Met Cys Ser Ala Arg Pro Pro Met Val Gln Thr Tyr Cys Thr
 20 25 30

Cys Leu Val Ser Pro Ala Leu Ala His Leu Pro Leu Tyr Ser Leu
 35 40 45

<210> 162

<211> 137

<212> PRT

<213> Homo sapien

<400> 162

Met Ser Gln Asp Thr Arg Ser Cys Pro Leu Phe Leu Ala Gln Leu Gly
 1 5 10 15

Arg Arg Lys Gly Leu Gln Ala Arg Ala Ala Gly Gln Ala Gly Leu Pro
 20 25 30

Leu Gly His Arg Thr Pro Leu Pro Pro Arg Pro Arg Leu His Ser His
 35 40 45

His His Lys Ala Gln Val Pro Ser His Trp Leu Pro Lys Lys Ala Thr
 50 55 60

Glu Arg Ile Phe Phe Leu Pro Leu Asn Val Ser Phe Pro Leu Gly Cys
 65 70 75 80

Leu Ser Val Ala Leu Pro Ser Gln Val Phe Leu Gly Met Leu Arg Ala
 85 90 95

Trp Arg Cys Thr Gly Gly Val Gln Trp His Leu Pro Pro Glu Leu Pro
 100 105 110

110

His Ser Leu Leu Arg Asn Leu Arg Gly Glu Gly Gly Ala Pro Gly Leu
115 120 125

Arg Glu Lys Glu Ser Val Leu Thr Phe
130 135

<210> 163
<211> 68
<212> PRT
<213> Homo sapien

<400> 163

Met Pro Val Ile Pro Ala Leu Trp Glu Ala Glu Ala Gly Gly Ser Arg
1 5 10 15

Gly Gln Glu Ile Glu Thr Ile Leu Ala Asn Ala Val Lys Ser Arg Leu
20 25 30

Leu Lys Ile Gln Lys Ile Ser Gln Ala Trp Trp Arg Ala Pro Val Val
35 40 45

Pro Ala Thr Arg Glu Ala Glu Ala Gly Glu Trp Leu Glu Pro Gly Arg
50 55 60

Arg Ser Leu Gln
65

<210> 164
<211> 17
<212> PRT
<213> Homo sapien

<400> 164

Met Asn Ile Thr Ser Tyr Lys Ser Phe Ser Ile Tyr Ile Asn Ala Ile
1 5 10 15

Asp

<210> 165
<211> 784
<212> PRT
<213> Homo sapien

<400> 165

Met Ala Ser Leu Asp Asp Pro Gly Glu Val Arg Glu Gly Phe Leu Cys
1 5 10 15

111

Pro Leu Cys Leu Lys Asp Leu Gln Ser Phe Tyr Gln Leu His Ser His
 20 25 30

Tyr Glu Glu Glu His Ser Gly Glu Asp Arg Asp Val Lys Gly Gln Ile
 35 40 45

Lys Ser Leu Val Gln Lys Ala Lys Lys Ala Lys Asp Arg Leu Leu Lys
 50 55 60

Arg Glu Gly Asp Asp Arg Ala Glu Ser Gly Thr Gln Gly Tyr Glu Ser
 65 70 75 80

Phe Ser Tyr Gly Gly Val Asp Pro Tyr Met Trp Glu Pro Gln Glu Leu
 85 90 95

Gly Ala Val Arg Ser His Leu Ser Asp Phe Lys Lys His Arg Ala Ala
 100 105 110

Arg Ile Asp His Tyr Val Val Glu Val Asn Lys Leu Ile Ile Arg Leu
 115 120 125

Glu Lys Leu Thr Ala Phe Asp Arg Thr Asn Thr Glu Ser Ala Lys Ile
 130 135 140

Arg Ala Ile Glu Lys Ser Val Val Pro Trp Val Asn Asp Gln Asp Val
 145 150 155 160

Pro Phe Cys Pro Asp Cys Gly Asn Lys Phe Ser Ile Arg Asn Arg Arg
 165 170 175

His His Cys Arg Leu Cys Gly Ser Ile Met Cys Lys Lys Cys Met Glu
 180 185 190

Leu Ile Ser Leu Pro Leu Ala Asn Lys Leu Thr Ser Ala Ser Lys Glu
 195 200 205

Ser Leu Ser Thr His Thr Ser Pro Ser Gln Ser Pro Asn Ser Val His
 210 215 220

Gly Ser Arg Arg Gly Ser Ile Ser Ser Met Ser Ser Val Ser Ser Val
 225 230 235 240

Leu Asp Glu Lys Asp Asp Asp Arg Ile Arg Cys Cys Thr His Cys Lys
 245 250 255

112

Asp Thr Leu Leu Lys Arg Glu Gln Gln Ile Asp Glu Lys Glu His Thr
 260 265 270

Pro Asp Ile Val Lys Leu Tyr Glu Lys Leu Arg Leu Cys Met Glu Lys
 275 280 285

Val Asp Gln Lys Ala Pro Glu Tyr Ile Arg Met Ala Ala Ser Leu Asn
 290 295 300

Ala Gly Glu Thr Thr Tyr Ser Leu Glu His Ala Ser Asp Leu Arg Val
 305 310 315 320

Glu Val Gln Lys Val Tyr Glu Leu Ile Asp Ala Leu Ser Lys Lys Ile
 325 330 335

Leu Thr Leu Gly Leu Asn Gln Asp Pro Pro Pro His Pro Ser Asn Leu
 340 345 350

Arg Leu Gln Arg Met Ile Arg Tyr Ser Ala Thr Leu Phe Val Gln Glu
 355 360 365

Lys Leu Leu Gly Leu Met Ser Leu Pro Thr Lys Glu Gln Phe Glu Glu
 370 375 380

Leu Lys Lys Lys Arg Lys Glu Glu Met Glu Arg Lys Arg Ala Val Glu
 385 390 395 400

Arg Gln Ala Ala Leu Glu Ser Gln Arg Arg Leu Glu Glu Arg Gln Ser
 405 410 415

Gly Leu Ala Ser Arg Ala Ala Asn Gly Glu Val Ala Ser Leu Arg Arg
 420 425 430

Gly Pro Ala Pro Leu Arg Lys Ala Glu Gly Trp Leu Pro Leu Ser Gly
 435 440 445

Gly Gln Gly Gln Ser Glu Asp Ser Asp Pro Leu Leu Gln Gln Ile His
 450 455 460

Asn Ile Thr Ser Phe Ile Arg Gln Ala Lys Ala Ala Gly Arg Met Asp
 465 470 475 480

Glu Val Arg Thr Leu Gln Glu Asn Leu Arg Gln Leu Gln Asp Glu Tyr
 485 490 495

113

Asp Gln Gln Gln Thr Glu Lys Ala Ile Glu Leu Ser Arg Arg Gln Ala
 500 505 510

Glu Glu Glu Asp Leu Gln Arg Glu Gln Leu Gln Met Leu Arg Glu Arg
 515 520 525

Glu Leu Glu Arg Glu Arg Glu Gln Phe Arg Val Ala Ser Leu His Thr
 530 535 540

Arg Thr Arg Ser Leu Asp Phe Arg Glu Ile Gly Pro Phe Gln Leu Glu
 545 550 555 560

Pro Ser Arg Glu Pro Arg Thr His Leu Ala Tyr Ala Leu Asp Leu Gly
 565 570 575

Ser Ser Pro Val Pro Ser Ser Thr Ala Pro Lys Thr Pro Ser Leu Ser
 580 585 590

Ser Thr Gln Pro Thr Arg Val Trp Ser Gly Pro Pro Ala Val Gly Gln
 595 600 605

Glu Arg Leu Pro Gln Ser Ser Met Pro Gln Gln His Glu Gly Pro Ser
 610 615 620

Leu Asn Pro Phe Asp Glu Glu Asp Leu Ser Ser Pro Met Glu Glu Ala
 625 630 635 640

Thr Thr Gly Pro Pro Ala Ala Gly Val Ser Leu Asp Pro Ser Ala Arg
 645 650 655

Ile Leu Lys Glu Tyr Asn Pro Phe Glu Glu Glu Asp Glu Glu Glu Glu
 660 665 670

Ala Val Ala Gly Asn Pro Phe Ile Gln Pro Asp Ser Pro Ala Pro Asn
 675 680 685

Pro Phe Ser Glu Glu Asp Glu His Pro Gln Gln Arg Leu Ser Ser Pro
 690 695 700

Leu Val Pro Gly Asn Pro Phe Glu Glu Pro Thr Cys Ile Asn Pro Phe
 705 710 715 720

Glu Met Asp Ser Asp Ser Gly Pro Glu Ala Glu Glu Pro Ile Glu Glu
 725 730 735

Glu Leu Leu Leu Gln Gln Ile Asp Asn Ile Lys Ala Tyr Ile Phe Asp

114

740

745

750

Ala Lys Gln Cys Gly Arg Leu Asp Glu Val Glu Val Leu Thr Glu Asn
 755 760 765

Leu Arg Glu Leu Lys His Thr Leu Ala Lys Gln Lys Gly Gly Thr Asp
 770 775 780

<210> 166
 <211> 96
 <212> PRT
 <213> Homo sapien

<400> 166

Met Ser Gly Thr Ser Thr Pro Ala Met Gly Val Cys Ala Glu Pro Leu
 1 5 10 15

Lys Val Asp Leu Ser Phe Gly Glu Pro Ser Glu Arg His Ser Trp Tyr
 20 25 30

Leu Trp Glu Trp Leu Trp Gly Val Leu Trp His His Thr Asp Asn Phe
 35 40 45

Val Phe Leu Ile Gly Ile His Gly Leu Gly Ser Trp Asp Gly Gly Arg
 50 55 60

Gly Lys Pro Gln Ser Pro Trp Lys Cys Met Gln Asn Ile Trp Val Cys
 65 70 75 80

Arg Cys Ile Ile Leu Arg Leu Trp Pro Leu Gln Phe Phe Gln Arg Val
 85 90 95

<210> 167
 <211> 215
 <212> PRT
 <213> Homo sapien

<400> 167

Met Ser Gly Glu Lys Arg Ser Pro Ser Val Asn Thr Gln Ala Pro Gly
 1 5 10 15

Thr Thr Arg Gly Ala Cys Lys Gly Thr Thr Ser Trp Glu Pro Ser Trp
 20 25 30

Pro His Cys Glu Glu Val Pro Gln Glu Glu Lys Phe Leu Arg Gly Thr
 35 40 45

115

Ser Ser Ala Gly Tyr Ser Pro Leu Gln His Gly Gly Pro Pro Ile Thr
 50 55 60

Pro Arg Lys Gly Gly Asn His Phe Leu Glu Lys Ser Lys Gly Ala Gln
 65 70 75 80

Thr Pro Ala Val Ser Leu Ser Pro Pro Gly Lys Phe Phe Phe Phe Phe
 85 90 95

Phe Phe Phe Gln Phe Leu Ala Ala Gly Gly Trp Gly Ser Ser Ser Val
 100 105 110

Val Ser Arg Arg Val Arg Pro Arg Gln Arg Pro Thr Ser Arg Arg Arg
 115 120 125

Pro Gly Val Glu Ser Ala Leu Asn Lys Ile Leu Ser Gln Leu Cys Val
 130 135 140

Pro Arg Arg Val Lys Gly Pro Val Pro Ala Gly Ala Arg Gly Ser Ala
 145 150 155 160

Trp Ser Trp Arg Pro Ala Arg Arg Leu Trp Arg Pro Leu Ser Glu Pro
 165 170 175

Arg Arg Pro Pro Arg Val Phe Val Asp Cys Arg Ser Pro Gly Arg Leu
 180 185 190

Gly Thr Val Thr Ala Pro Lys Ala Gly Asp Val Ala Ala Leu Gly Arg
 195 200 205

Arg Leu Leu Pro Pro Pro Leu
 210 215

<210> 168

<211> 67

<212> PRT

<213> Homo sapien

<400> 168

Met Ser Leu Arg Pro Leu Met Gly Leu Asp Pro Ala Ser Gly Lys Leu
 1 5 10 15

Leu Ser His Thr Leu Pro His Leu Thr Ile His Val Ser Phe Leu Pro
 20 25 30

Leu Ser Leu His Ala Cys His Ser His Ser Leu Cys Met Pro Ala Phe
 35 40 45

116

Pro Ala Ala Pro Cys Thr Leu Arg Leu Trp Pro Phe Pro Ala Phe Cys
 50 55 60

Leu Pro Gly
 65

<210> 169
 <211> 102
 <212> PRT
 <213> Homo sapien

<400> 169

Met Pro Thr Pro Ser Leu Ser Tyr Thr Gly Glu Gly Gln Lys Gln Asp
 1 5 10 15

Val Glu Phe Ala Trp Pro Trp Pro Thr Gln Thr Thr Ser Pro Ser Ser
 20 25 30

Gly Ser Gln Ala Arg Glu Lys Gly Ile Ser Arg Thr Thr Val Pro Gly
 35 40 45

Ile Lys His Lys Leu Lys Leu Ala Gly Ala Gly Met Val Ser Gln Thr
 50 55 60

Ser Glu Leu Pro Pro His Leu His Leu Phe Gly Leu Thr Gly Asn Met
 65 70 75 80

Ala Ser Leu Gly Pro Ala Gly Arg Ala His Gly Arg Arg Leu Leu Ser
 85 90 95

His Gln Ala Thr Glu Asp
 100

<210> 170
 <211> 38
 <212> PRT
 <213> Homo sapien

<400> 170

Met Arg Trp Val Glu Ser Thr Leu Glu Asn Ile Ile His Lys Ala Asn
 1 5 10 15

Asn Gln Thr Gly Arg Lys Asp Asn Thr Lys Ser Leu Thr Ser Thr Ile
 20 25 30

His Gly Ser Leu Ala Leu

117

35

<210> 171
 <211> 52
 <212> PRT
 <213> Homo sapien

<400> 171

Met Ser Gln Phe Leu Asp Leu Leu Thr Val Ile Asn Ser Gln Ala His
 1 5 10 15

Ser Leu Ser Arg Val Asn Asp Thr Thr Phe Asn Leu Glu Lys Ser Ser
 20 25 30

Ile Thr Phe Tyr His Phe Gln Ile Ser Phe Leu Pro Ser Leu His Thr
 35 40 45

Lys Asp Ser Ser
 50

<210> 172
 <211> 45
 <212> PRT
 <213> Homo sapien

<400> 172

Met Lys Leu His Met Thr Arg Phe Ser Phe Lys Leu Ile Leu Lys Cys
 1 5 10 15

Leu Leu Leu Leu Gly Val Leu Asn Ser Phe Thr Val Leu Val Thr Thr
 20 25 30

Arg Leu Ile Ser Ala Ile Leu Tyr Tyr Ala Ile Phe Ser
 35 40 45

<210> 173
 <211> 34
 <212> PRT
 <213> Homo sapien

<400> 173

Met Leu Thr Cys Cys Leu Asn Thr Glu Asp Ala Ile Leu Phe Ser Arg
 1 5 10 15

Ala Leu Phe Leu His Gly Lys Ile Val Asp Pro Thr Ile Lys Cys Gly
 20 25 30

Tyr Ser

118

<210> 174
<211> 29
<212> PRT
<213> Homo sapien

<400> 174

Met Phe Leu Arg Val Ala Gly Asn Ile Leu Leu Glu Asn Gln Met Asn
1 5 10 15

Thr Gly Arg Tyr Gly Tyr Gln Glu Leu Arg Leu Ile Leu
20 25

<210> 175
<211> 21
<212> PRT
<213> Homo sapien

<400> 175

Met Asn Ile His Asn Pro Ser Ser Phe Leu Val Ser Leu Cys Asn Leu
1 5 10 15

Ala Ser Gln Asn Tyr
20

<210> 176
<211> 78
<212> PRT
<213> Homo sapien

<400> 176

Met Lys Ser Lys Ser Ser Leu Pro Ser Leu Gly Val Lys Leu His Asn
1 5 10 15

Leu Ile Ser Ala Ser Ser Cys Cys Leu Tyr Phe Arg Lys Lys Lys Gln
20 25 30

Thr Val Gly Phe Pro Cys Glu Ser Asp Phe Ser Leu Ser Ala Tyr Cys
35 40 45

Ala Ser Ala Phe Pro Cys Val Ser His Asp Leu Met Ala Ser Leu Thr
50 55 60

Pro Asn Cys Met Tyr Pro Ala Gln Ile Ser Ile Leu Pro Gln
65 70 75

<210> 177

119

<211> 37
<212> PRT
<213> Homo sapien

<400> 177

Met Ser Leu Lys Arg Tyr Cys Leu Ala Leu Lys Ser Ala Ser Val Val
1 5 10 15

Lys Val Leu Gln Tyr Ile His Thr Glu Glu Lys Cys Gly Ser Glu Ser
20 25 30

Pro Trp Ala Ala Phe
35

<210> 178
<211> 21
<212> PRT
<213> Homo sapien

<400> 178

Met Thr Lys Pro Thr Arg Leu Leu Pro Lys Asn Leu Ile Tyr Cys Pro
1 5 10 15

Arg Ser Arg Leu Ser
20

<210> 179
<211> 57
<212> PRT
<213> Homo sapien

<400> 179

Met Cys Cys Val Ser Gly Gly Trp Lys Lys Leu Ile Gln Leu Trp Pro
1 5 10 15

Ser Ser Ile Met Val Arg Tyr His Tyr His Leu Gln Ile Ser Ser Lys
20 25 30

Ile Gly Gly Leu Pro Ser Trp Ala Leu Arg Ala Asn Ser Leu Gly Gln
35 40 45

Phe Phe Leu Leu His Ser Thr Ile Glu
50 55

<210> 180
<211> 23
<212> PRT
<213> Homo sapien

120

<400> 180

Met Asn Phe Ile Leu Cys Pro Cys Phe Asp Ala Ser Thr Gln Phe Leu
1 5 10 15

Val Ile Ser Val Lys Tyr Asn
20

<210> 181

<211> 92

<212> PRT

<213> Homo sapien

<400> 181

Met Arg Lys Lys Cys Arg Leu Thr Ser Pro Cys Trp Trp Arg Trp Arg
1 5 10 15

Arg Gly Gly Glu Gln Leu Leu Arg Ile Phe Val Leu Ile Phe Ser Cys
20 25 30

Ser Pro Pro Asp Asn Thr Val Gly Thr Cys Ile Leu Trp Gly Ile Phe
35 40 45

Ser Gly Lys Ser Lys Asp Cys Glu Trp Leu His Leu Ile Ser Thr Leu
50 55 60

Arg Asn Ala Glu Ala Cys Ser Ser Gln Val Leu Lys Asn Tyr Leu Glu
65 70 75 80

Lys Lys Asn Arg Pro Ile Lys Ser Thr Val Lys Arg
85 90

<210> 182

<211> 27

<212> PRT

<213> Homo sapien

<400> 182

Met Leu His Pro Thr Glu Met Gly Pro Gln Val Pro Asp Leu Pro Phe
1 5 10 15

Arg Arg Gly Arg Gln Val Gly Glu Gln Gln Met
20 25

<210> 183

<211> 166

<212> PRT

<213> Homo sapien

121

<400> 183

Met Pro Leu Thr Pro Glu Pro Pro Ser Gly Arg Val Glu Gly Pro Pro
 1 5 10 15

Ala Trp Glu Ala Ala Pro Trp Pro Ser Leu Pro Cys Gly Pro Cys Ile
 20 25 30

Pro Ile Met Leu Val Leu Ala Thr Leu Ala Ala Leu Phe Ile Leu Thr
 35 40 45

Thr Ala Val Leu Ala Glu Arg Leu Phe Arg Arg Ala Leu Arg Pro Asp
 50 55 60

Pro Ser His Arg Ala Pro Thr Leu Val Trp Arg Pro Gly Gly Glu Leu
 65 70 75 80

Trp Ile Glu Pro Met Gly Thr Ala Arg Glu Arg Ser Glu Asp Trp Tyr
 85 90 95

Gly Ser Ala Val Pro Leu Leu Thr Asp Arg Ala Pro Glu Pro Pro Thr
 100 105 110

Gln Val Gly Thr Leu Glu Ala Arg Ala Thr Ala Pro Pro Ala Pro Ser
 115 120 125

Ala Pro Asn Ser Ala Pro Ser Asn Leu Gly Pro Gln Thr Val Leu Glu
 130 135 140

Val Pro Ala Arg Ser Thr Phe Trp Gly Pro Gln Pro Trp Glu Gly Arg
 145 150 155 160

Pro Pro Pro Gln Ala Trp
 165

<210> 184

<211> 80

<212> PRT

<213> Homo sapien

<400> 184

Met Leu Thr Ser His Phe Ile Leu Ile Pro Val Ile Phe Ser Leu Gln
 1 5 10 15

Tyr Gln Cys Leu Gly Ala Arg Lys Leu Cys Gln Cys Gln Trp Leu Trp
 20 25 30

122

Arg Trp Gln Lys Lys Gly Gly Gln Pro Pro Gly Thr Ala Glu Ser Lys
 35 40 45

Pro Asp Ser Gln Pro Gln Lys Val Gly Gln Asp Ala Ala Asn Ser Ser
 50 55 60

Asn Pro Lys Lys Ala Ala Glu Ile Thr Val Ile Gln Gln Thr Tyr Phe
 65 70 75 80

<210> 185

<211> 159

<212> PRT

<213> Homo sapien

<400> 185

Met Asp Thr Ile Leu Val Phe Ser Leu Ile Ile Ala Ser Tyr Asp Ala
 1 5 10 15

Asn Lys Lys Asp Leu Arg Asp Ser Ser Cys Arg Leu Glu Gln Leu Pro
 20 25 30

Gly Ile Phe Pro Lys Asp Val Arg Ser Ile Arg Glu Leu Gln Met Gln
 35 40 45

Glu Thr His Thr Glu Thr Lys Arg Thr Thr Phe Ile Gln Asn Arg Thr
 50 55 60

Ile Ala Thr Leu Gln Cys Leu Gly Ser Asp Ser Lys Val Lys Val Asn
 65 70 75 80

Leu Val Tyr Leu Glu Arg Arg Pro Lys Val Lys His Ile Leu Lys Asn
 85 90 95

Leu Arg Ile Ile Ala Ala Pro Arg Arg Asn Ser Ser Ala Ser Ser Ser
 100 105 110

Cys His Leu Ile Pro Thr Ser Lys Phe Gln Thr Gly Ser Leu Leu Lys
 115 120 125

Gly Lys Val Ser Met Pro Arg Ser Gln Glu Ala Val Pro Met Pro Val
 130 135 140

Val Val Glu Met Ala Lys Glu Gly Arg Pro Ala Thr Trp Asp Ser
 145 150 155

<210> 186

<211> 928

123

<212> PRT

<213> Homo sapien

<400> 186

Met Ala Glu Gly Lys Glu Lys Gln Val Thr Ser Tyr Met Asp Gly Ser
 1 5 10 15

Arg Pro Tyr Asp Val Ser Met Thr Tyr Ile His Lys Ala Gly Gly Pro
 20 25 30

Asp Gln Gln Glu Leu Val Met Leu Thr Cys Thr Val Pro Leu Asp Ser
 35 40 45

Cys Cys His Leu Pro Gln Ala Arg Thr Asn Tyr Arg Lys Tyr Phe Arg
 50 55 60

Ser Glu Ala Ala Phe Thr Leu Ala Asp Phe Ile Tyr Lys Ser Met Ile
 65 70 75 80

Arg Val Asn Ser Ser Arg Leu Val Arg Val Thr Gln Val Glu Asn Glu
 85 90 95

Glu Lys Leu Lys Glu Leu Glu Gln Phe Ser Ile Trp Asn Phe Phe Ser
 100 105 110

Ser Phe Leu Lys Glu Lys Leu Asn Asp Thr Tyr Val Asn Val Gly Leu
 115 120 125

Tyr Ser Thr Lys Thr Cys Leu Lys Val Glu Ile Ile Glu Lys Asp Thr
 130 135 140

Lys Tyr Ser Val Ile Val Ile Arg Arg Ser Trp Asp Val Ile Arg Val
 145 150 155 160

Asn Ser Ser Arg Leu Val Arg Val Thr Gln Val Glu Asn Glu Glu Lys
 165 170 175

Leu Lys Glu Leu Glu Gln Phe Ser Ile Trp Asn Phe Phe Ser Ser Phe
 180 185 190

Leu Lys Glu Lys Leu Asn Asp Thr Tyr Val Asn Val Gly Leu Tyr Ser
 195 200 205

Thr Lys Thr Cys Leu Lys Val Glu Ile Ile Glu Lys Asp Thr Lys Tyr
 210 215 220

124

Ser Val Ile Val Ile Arg Arg Ser Trp Asp Val Ile Arg Val Asn Ser
 225 230 235 240

Ser Arg Leu Val Arg Val Thr Gln Val Glu Asn Glu Glu Lys Leu Lys
 245 250 255

Glu Leu Glu Gln Phe Ser Ile Trp Asn Phe Phe Ser Ser Phe Leu Lys
 260 265 270

Glu Lys Leu Asn Asp Thr Tyr Val Asn Gly Ile Pro Trp Thr Lys Val
 275 280 285

Asp Tyr Phe Asp Asn Gly Ile Ile Cys Lys Leu Ile Glu His Asn Gln
 290 295 300

Arg Gly Ile Leu Ala Met Leu Asp Glu Glu Cys Leu Arg Pro Gly Val
 305 310 315 320

Val Ser Asp Ser Thr Phe Leu Ala Lys Leu Asn Gln Leu Phe Ser Lys
 325 330 335

His Gly His Tyr Glu Ser Lys Val Thr Gln Asn Ala Gln Arg Gln Tyr
 340 345 350

Asp His Thr Met Gly Leu Ser Cys Phe Arg Ile Cys His Tyr Ala Gly
 355 360 365

Lys Val Thr Tyr Asn Val Thr Ser Phe Ile Asp Lys Asn Asn Asp Leu
 370 375 380

Leu Phe Arg Asp Leu Leu Gln Ala Met Trp Lys Ala Gln His Pro Leu
 385 390 395 400

Leu Arg Ser Leu Phe Pro Glu Gly Asn Pro Lys Gln Ala Ser Leu Lys
 405 410 415

Arg Pro Pro Thr Ala Gly Ala Gln Phe Lys Ser Ser Val Ala Ile Leu
 420 425 430

Met Lys Asn Leu Tyr Ser Lys Ser Pro Asn Tyr Ile Arg Cys Ile Lys
 435 440 445

Pro Asn Glu His Gln Gln Arg Gly Gln Phe Ser Ser Asp Leu Val Ala
 450 455 460

Thr Gln Ala Arg Tyr Leu Gly Leu Leu Glu Asn Val Arg Val Arg Arg

125

| | | | | | | |
|---|---|-----|-----|-----|-----|-----|
| 465 | | 470 | | 475 | | 480 |
| Ala Gly Tyr | Ala His Arg Gln Gly Tyr Gly Pro Phe Leu Glu Arg Tyr | | | | | |
| | 485 | | 490 | | 495 | |
| Arg Leu Leu Ser Arg Ser Thr Trp Pro His Trp Asn Gly Gly Asp Arg | | | | | | |
| | 500 | | 505 | | 510 | |
| Glu Gly Val Glu Lys Val Leu Gly Glu Leu Ser Met Ser Ser Gly Glu | | | | | | |
| | 515 | | 520 | | 525 | |
| Leu Ala Phe Gly Lys Thr Lys Ile Phe Ile Arg Ser Pro Lys Thr Leu | | | | | | |
| | 530 | | 535 | | 540 | |
| Phe Tyr Leu Glu Glu Gln Arg Arg Leu Arg Leu Gln Gln Leu Ala Thr | | | | | | |
| | 545 | | 550 | | 555 | 560 |
| Leu Ile Gln Lys Ile Tyr Arg Gly Trp Arg Cys Arg Thr His Tyr Gln | | | | | | |
| | 565 | | 570 | | 575 | |
| Leu Met Arg Lys Ser Gln Ile Leu Ile Ser Ser Trp Phe Arg Gly Asn | | | | | | |
| | 580 | | 585 | | 590 | |
| Met Ala Arg Lys Asn Tyr Arg Lys Tyr Phe Arg Ser Glu Ala Ala Leu | | | | | | |
| | 595 | | 600 | | 605 | |
| Thr Leu Ala Asp Phe Ile Tyr Lys Ser Met Val Gln Lys Phe Leu Leu | | | | | | |
| | 610 | | 615 | | 620 | |
| Gly Leu Lys Asn Asn Leu Pro Ser Thr Asn Val Leu Asp Lys Thr Trp | | | | | | |
| | 625 | | 630 | | 635 | 640 |
| Pro Ala Ala Pro Tyr Lys Cys Leu Ser Thr Ala Asn Gln Glu Leu Gln | | | | | | |
| | 645 | | 650 | | 655 | |
| Gln Leu Phe Tyr Gln Trp Lys Ala Thr Pro Val Pro Pro Ser Ser Gln | | | | | | |
| | 660 | | 665 | | 670 | |
| Cys Lys Arg Phe Arg Asp Gln Leu Ser Pro Lys Gln Val Glu Ile Leu | | | | | | |
| | 675 | | 680 | | 685 | |
| Arg Glu Lys Leu Cys Ala Ser Glu Leu Phe Lys Gly Lys Lys Ala Ser | | | | | | |
| | 690 | | 695 | | 700 | |
| Tyr Pro Gln Ser Val Pro Ile Pro Phe Cys Gly Asp Tyr Ile Gly Leu | | | | | | |
| | 705 | | 710 | | 715 | 720 |

126

Gln Gly Asn Pro Lys Leu Gln Lys Leu Lys Gly Gly Glu Glu Gly Pro
 725 730 735

Val Leu Met Ala Glu Ala Val Lys Lys Val Asn Arg Gly Asn Gly Lys
 740 745 750

Thr Ser Ser Arg Ile Leu Leu Leu Thr Lys Gly His Val Ile Leu Thr
 755 760 765

Asp Thr Lys Lys Ser Gln Ala Lys Ile Val Ile Gly Leu Asp Asn Val
 770 775 780

Ala Gly Val Ser Val Thr Ser Leu Lys Asp Gly Leu Phe Ser Leu His
 785 790 795 800

Leu Ser Glu Met Ser Ser Val Gly Ser Lys Gly Asp Phe Leu Leu Val
 805 810 815

Ser Glu His Val Ile Glu Leu Leu Thr Lys Met Tyr Arg Ala Val Leu
 820 825 830

Asp Ala Thr Gln Arg Gln Leu Thr Val Thr Val Thr Glu Lys Phe Ser
 835 840 845

Val Arg Phe Lys Glu Asn Ser Val Ala Val Lys Val Val Gln Gly Pro
 850 855 860

Ala Gly Gly Asp Asn Ser Lys Leu Arg Tyr Lys Lys Lys Gly Ser His
 865 870 875 880

Cys Leu Glu Val Thr Val Gln Gln Leu Thr Ala Gly Tyr His Ala Gly
 885 890 895

Gln Gly Glu Leu Ile Asn Phe Ser Ser Cys Leu Gln Ile Asn Leu Leu
 900 905 910

Ser Glu His Lys Pro Arg Ala Ser Gly Thr Pro Cys Phe Glu Leu Arg
 915 920 925

<210> 187

<211> 96

<212> PRT

<213> Homo sapien

<400> 187

127

Met Arg Arg Cys Tyr Ser Ile Pro Val Cys Lys Cys Ala Gly Met Pro
 1 5 10 15

Ala Leu Ser Asp Gly Gly His Asp Asn Met Ala His Ala Phe Lys Leu
 20 25 30

Thr Ser Asn Cys Phe Trp Thr Thr Phe Asn Arg Gly Ser His Tyr His
 35 40 45

Gly Phe Lys Glu Pro Cys Gln Pro Arg Lys His Leu Thr Ala Gly Thr
 50 55 60

Ala Gly Trp Ser Cys Cys Trp Leu Glu Val Tyr Ala Arg Ile Ala Lys
 65 70 75 80

Asp Ser Trp Arg Met Gly Ser Pro Tyr Leu Cys Arg Leu Ala Ala Leu
 85 90 95

<210> 188
 <211> 47
 <212> PRT
 <213> Homo sapien

<400> 188

Met Tyr Gln Lys Asp Leu Tyr Ile Ser Gln Arg Gly Thr Gln Ala Lys
 1 5 10 15

Leu Lys Ile Tyr Lys His Asn Gln Phe Thr Arg Glu Ile Ile Leu Thr
 20 25 30

Val Phe Phe Leu Phe Phe Gln Thr Leu Leu Phe His Gly Lys Lys
 35 40 45

<210> 189
 <211> 644
 <212> PRT
 <213> Homo sapien

<400> 189

Met Met Ile Ile Ala Leu Glu Phe Pro His Leu Val Val Asp Leu Ala
 1 5 10 15

Asp Asn Asn Trp Gln Cys Asp Asp Ser Val Ala Val Phe Gln Asn Phe
 20 25 30

Ile Ser Glu Ser Trp Arg Lys Lys Trp Asn Val Ile Cys Asn Arg Ser
 35 40 45

128

| | | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Ile | Gly | Ser | Glu | Glu | Ala | Asn | Gly | Gly | Thr | Pro | Gln | Ser | Arg | Ile | Ser | 50 | 55 | 60 |
| Arg | Glu | Thr | Arg | Leu | Pro | Pro | Ile | His | Leu | His | Arg | Met | Lys | Ser | Leu | 65 | 70 | 75 |
| Ile | Arg | Ser | Lys | Ala | Glu | Arg | Pro | Gln | Gly | Gly | Arg | His | Thr | Gly | Ile | 85 | 90 | 95 |
| Ser | Thr | Leu | Gly | Lys | Lys | Ala | Lys | Ala | Gly | Ser | Gly | Leu | Arg | Lys | Lys | 100 | 105 | 110 |
| Gln | Arg | Arg | Leu | Pro | Arg | Ser | Val | Arg | Ser | Thr | Arg | Asp | Val | Gln | Ala | 115 | 120 | 125 |
| Ala | Gly | Lys | Lys | Glu | Asp | Ala | Pro | Gln | Asp | Leu | Ala | Leu | Ala | Val | Cys | 130 | 135 | 140 |
| Leu | Ser | Val | Phe | Ile | Thr | Phe | Leu | Val | Ala | Phe | Ser | Leu | Gly | Ala | Phe | 145 | 150 | 155 |
| Thr | Arg | Pro | Tyr | Val | Asp | Arg | Leu | Trp | Gln | Lys | Lys | Cys | Gln | Ser | Lys | 165 | 170 | 175 |
| Ser | Pro | Gly | Leu | Asp | Asn | Ala | Tyr | Ser | Asn | Glu | Gly | Phe | Tyr | Asp | Asp | 180 | 185 | 190 |
| Met | Glu | Ala | Ala | Gly | His | Thr | Pro | His | Pro | Glu | Thr | His | Leu | Arg | Gln | 195 | 200 | 205 |
| Val | Phe | Pro | His | Leu | Ser | Leu | Tyr | Glu | Asn | Gln | Thr | Pro | Phe | Trp | Val | 210 | 215 | 220 |
| Thr | Gln | Pro | His | Pro | His | Ala | Thr | Val | Ile | Pro | Asp | Arg | Thr | Leu | Gly | 225 | 230 | 235 |
| Arg | Ser | Arg | Lys | Asp | Pro | Gly | Ser | Ser | Gln | Ser | Pro | Gly | Gln | Cys | Gly | 245 | 250 | 255 |
| Asp | Asn | Thr | Gly | Ala | Gly | Ser | Gly | Asn | Asp | Gly | Ala | Val | Tyr | Ser | Ile | 260 | 265 | 270 |
| Leu | Gln | Arg | His | Pro | His | Ala | Gly | Asn | Arg | Glu | Leu | Met | Ser | Ala | Ala | 275 | 280 | 285 |

129

Gln Asp His Ile His Arg Asn Asp Ile Leu Gly Glu Trp Thr Tyr Glu
 290 295 300

Thr Val Ala Gln Glu Glu Pro Leu Ser Ala His Ser Val Gly Val Ser
 305 310 315 320

Ser Val Ala Gly Thr Ser His Ala Val Ser Gly Ser Ser Arg Tyr Asp
 325 330 335

Ser Asn Glu Leu Asp Leu Pro Leu Ser Gly Glu Ile Thr Ala Ser Leu
 340 345 350

Cys Lys Met Leu Thr His Ala Glu Ala Gln Arg Thr Gly Asp Ser Lys
 355 360 365

Glu Arg Gly Gly Thr Glu Gln Ser Leu Trp Asp Ser Gln Met Glu Phe
 370 375 380

Ser Lys Glu Arg Gln Val Ser Ser Ser Ile Asp Leu Leu Ser Ile Gln
 385 390 395 400

Gln Pro Arg Leu Ser Gly Ala Arg Ala Glu Glu Ala Leu Ser Ala His
 405 410 415

Tyr Ser Glu Val Pro Tyr Gly Asp Pro Arg Asp Thr Gly Pro Ser Val
 420 425 430

Phe Pro Pro Arg Trp Asp Ser Gly Leu Asp Val Thr Pro Ala Asn Lys
 435 440 445

Glu Pro Val Gln Lys Ser Thr Pro Ser Asp Thr Cys Cys Glu Leu Glu
 450 455 460

Ser Asp Cys Asp Ser Asp Glu Gly Ser Leu Phe Thr Leu Ser Ser Ile
 465 470 475 480

Ser Ser Glu Ser Ala Arg Ser Lys Thr Glu Glu Ala Val Pro Asp Glu
 485 490 495

Glu Ser Leu Gln Asp Glu Ser Ser Gly Ala Ser Lys Asp Asn Val Thr
 500 505 510

Ala Val Asp Ser Leu Glu Glu Asn Val Thr Phe Gln Thr Ile Pro Gly
 515 520 525

Lys Cys Lys Asn Gln Glu Asp Pro Phe Glu Lys Pro Leu Ile Ser Ala

130

530

535

540

Pro Asp Ser Gly Met Tyr Lys Thr His Leu Glu Asn Ala Ser Asp Thr
 545 550 555 560

Asp Arg Ser Glu Gly Leu Ser Pro Trp Pro Arg Ser Pro Gly Asn Ser
 565 570 575

Pro Leu Gly Asp Glu Phe Pro Gly Met Phe Thr Tyr Asp Tyr Asp Thr
 580 585 590

Ala Leu Gln Ser Lys Ala Ala Glu Trp His Cys Ser Leu Arg Asp Leu
 595 600 605

Glu Phe Ser Asn Val Asp Val Leu Gln Gln Thr Pro Pro Cys Ser Ala
 610 615 620

Glu Val Pro Ser Asp Pro Asp Lys Ala Ala Phe His Glu Arg Asp Ser
 625 630 635 640

Asp Ile Leu Lys

<210> 190

<211> 48

<212> PRT

<213> Homo sapien

<400> 190

Met Trp Thr Phe Tyr Ser Lys His His His Val Leu Leu Lys Phe Pro
 1 5 10 15

Gln Ile Leu Val Asp Val Leu Gln Gln Thr Pro Pro Cys Ser Ala Glu
 20 25 30

Val Pro Ser Asp Pro Asp Lys Ala Ala Phe His Glu Arg Phe Leu Phe
 35 40 45